



VNIVERSITAT DE VALÈNCIA

Programa de Doctorado en Economía Industrial

**Aproximaciones metodológicas para la obtención
de bases de datos de calidad. Instrumentos de
análisis del cambio económico y social.**

Tesis doctoral de

Virgilio Pérez Giménez

Dirigida por

José M. Pavía Miralles

Cristina Aybar Arias

Noviembre de 2022

A Laura, Álvaro e Inés

Índice

Resumen	6
Introducción.....	6
Contexto	6
Motivación	7
Objetivos.....	7
Metodología	8
Datos censales	8
Datos obtenidos de encuestas (métodos probabilísticos)	10
Datos obtenidos de encuestas (métodos no probabilísticos)	12
Conclusiones y futuros trabajos	13
Bibliografía	15
Anexo I	18
Anexo II.....	31
Anexo III.....	44
Anexo IV	74
Anexo V.....	89
Anexo VI	106

Resumen

Los continuos avances tecnológicos junto con el aumento y generalización del uso de Internet, entre otros factores, está facilitando que los datos generados por el conjunto de la sociedad puedan ser recolectados por cualquier agente. No obstante, estos datos generados de forma masiva requieren de un tratamiento adecuado antes de ser analizados y utilizados en investigación. De hecho, se considera que los investigadores dedican en torno a un 80% de su tiempo a preparar los datos, y el 20% restante a analizarlos. En la presente investigación se ofrecen técnicas y herramientas que permiten reducir considerablemente el tiempo y el esfuerzo que el investigador dedica a esa fase inicial y se muestran ejemplos de uso mediante su utilización para responder preguntas de investigación sustantivas. También se proponen algunos procedimientos para que la generación de bases de datos sea de calidad, considerándose una mejora sustantiva la puesta a disposición de los datos a otros investigadores y a la sociedad en general. Para ello se han manejado datos de diferentes tipos y provenientes de diferentes fuentes, mostrando la versatilidad de los métodos y herramientas propuestas. Al trabajo realizado se le ha dado visibilidad en varios congresos nacionales e internacionales, y mediante diversas publicaciones científicas, las cuales se comentan en profundidad a lo largo del presente documento, quedando disponibles para su lectura y análisis en los Anexos I a VI.

Introducción

Contexto

Los datos son la gran revolución de nuestro tiempo. Cualquiera de nuestros conciudadanos está acostumbrado a escuchar y a utilizar en sus conversaciones diarias términos como Big data, Open data, Data science o Social data, teniendo todos ellos un punto en común: la obtención de información útil que permita tomar las decisiones adecuadas (Curto-Rodríguez, 2021; Guillén et al., 2015). Muchos de estos datos son ofrecidos libre y gratuitamente en repositorios web y en los portales de multitud de instituciones públicas, con el doble objetivo de generar transparencia, confianza y conocimiento y permitir que la sociedad sea más eficiente. No obstante, de poco o nada sirve tener acceso a los datos si no sabemos qué hacer con ellos ni cómo explotarlos (Rivas-Rebaque et al., 2019).

En los últimos años han aumentado considerablemente las iniciativas relacionadas con ofrecer los datos en abierto (*open data*) y la reutilización de la información (Peset et al., 2017). Lamentablemente, ofrecer datos en abierto no es garantía suficiente para que otros agentes puedan utilizar tales datos en sus estudios, análisis o investigaciones. ¿Podemos fiarnos de los datos disponibles en un determinado portal? Si bien es cierto que desinformar no es uno de los objetivos (a priori) de las instituciones públicas en países desarrollados, los errores son humanos, y existe la posibilidad de que existan errores en tales datos, siendo necesario, recomendable cuanto menos, implementar técnicas y procedimientos de comprobación/verificación (Abella et al., 2018).

Otro aspecto por el que el investigador debe preocuparse es por el tipo de fichero (formato) en el que se presentan los datos que desea obtener. Asegurar un adecuado nivel de interoperabilidad resulta indispensable en la reutilización de los datos, es decir, es fundamental que los datos que se ofrecen de manera abierta puedan ser fácilmente reutilizados por terceros (Poisot et al., 2019). En la actualidad existe una amplia variedad de formatos, debido en gran medida a las múltiples aplicaciones informáticas

disponibles para el tratamiento y análisis de los datos (Verna & Gupta, 2012). En este sentido, otro reto del investigador es conocer qué software es capaz de leer/importar los datos obtenidos, además de conocer el funcionamiento de dicho software. Si bien es cierto que en la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información (documento que contiene las pautas básicas para la reutilización de documentos y recursos de información elaborados o custodiados por el sector público español) se indican algunas medidas que contribuyen a que los datos abiertos cumplan unos mínimos de interoperabilidad y homogeneidad (BOE, 2013), la realidad con la que se encuentra el investigador es que debe manejar con fluidez diversas aplicaciones informáticas, circunstancia que conlleva una menor especialización en su área de conocimiento, además del desgaste que conlleva adaptar los datos al formato deseado.

Motivación

La formulación de hipótesis y teorías sobre cómo funciona el mundo y su contrastación empírica constituyen la base del método científico y del progreso tecnológico y social. Una vez se dispone de una teoría o hipótesis, la labor del/la investigador/a es tratar de evaluar la compatibilidad de esta con la realidad, contrastando su verosimilitud y su capacidad para explicar los hechos que observamos. Para ello, el/la investigador/a trata de derivar tantas consecuencias contrastables (o potencialmente observables) como sea posible, para una vez que se dispone de los datos adecuados estudiar la congruencia entre teoría y realidad, empleando las metodologías y métodos (estadísticos) apropiados. Disponer de datos adecuados y de la suficiente calidad, por tanto, es la base del progreso y el avance científico.

Los avances tecnológicos permiten, por un lado, realizar acciones muy complicadas, incluso impensables, hace unos años y, por otro lado, obtener (acceder a) datos con relativa facilidad. Estos avances, junto con el enorme abanico de áreas y campos de investigación existente actualmente, permiten que los investigadores puedan plantearse y/o abordar cuestiones muy concretas, ofreciendo posibles soluciones a problemas complejos. No obstante, en muchas ocasiones los investigadores se encuentran (nos encontramos) con enormes dificultades para obtener datos de calidad relacionados con la cuestión a abordar, que permitan elaborar trabajos académicos relevantes y responder a preguntas de relevancia social con consistencia. Esta reflexión nos lleva a diferenciar la labor del investigador en dos fases: i) la obtención de los datos y su preparación para que estos sean de calidad, y ii) el análisis de los datos obtenidos, fase en la que el investigador está especializado (Riquelme et al., 2006).

En este sentido, la principal motivación que nos ha llevado a realizar este trabajo ha sido la de facilitar a la comunidad científica técnicas y herramientas que permitan a los investigadores centrarse en el campo en el que están especializados, reduciendo/resolviendo los problemas que se suelen encontrar en la fase de obtención y preparación de datos.

Objetivos

El presente trabajo tiene por objeto proponer métodos y soluciones para la generación de datos de calidad. Sin perjuicio de que las metodologías desarrolladas y los datos generados puedan ser utilizados por otra/os investigadore/as para responder sus propias cuestiones, esta Tesis doctoral tiene como

principal objetivo generar (bases de) datos de calidad, que permitan responder a algunas de las preguntas que actualmente ocupan al Grupo de Investigación en Procesos Electorales y Opinión Pública (GIPEyOP), en el que el autor del presente documento está integrado. La idea fuerza de este trabajo es desarrollar herramientas metodológicas y conjuntos de datos provenientes de entornos naturales que permitan encontrar respuestas a algunas preguntas. En particular, aproximaciones metodológicas que permitan construir conjuntos de datos de calidad que ayuden a avanzar en la comprensión de cómo ha ido cambiando la sociedad española y mediante qué mecanismos la ciudadanía ha ido modificando sus escalas de valores y preferencias.

Cualquier investigación y/o avance científico se sustenta en los trabajos que previamente han desarrollado otros investigadores. Del mismo modo que nos subimos *a hombros de gigantes* para seguir creciendo, la aportación de esta Tesis doctoral está fundamentada en ayudar a los investigadores a resolver algunos de los problemas con los que se encuentran de forma recurrente. En este sentido, los objetivos que pretende abordar esta Tesis doctoral son:

- Proponer métodos, herramientas y/o procedimientos que sirvan a otros investigadores a automatizar procesos en la fase de obtención y (pre)procesamiento de datos.
- Proponer métodos y soluciones para la generación de (bases de) datos de calidad.
- Mostrar los resultados conseguidos tras aplicar las medidas propuestas, comentadas en los dos puntos anteriores, que han dado lugar a varias publicaciones científicas.
- Visibilizar las ventajas de disponer de (bases de) datos de calidad de forma libre y gratuita.

Metodología

En una sociedad abierta, moderna y transparente, el acceso a información detallada y confiable es fundamental. Los datos básicos no solo deben ser accesibles, sino que también deben estar disponibles en formatos fáciles de usar. En esta investigación se presentan diferentes técnicas y estrategias que nos han permitido obtener y generar conjuntos de datos de calidad, los cuales se han puesto a disposición de la comunidad científica y de la sociedad en general (en repositorios web y como material suplementario de las publicaciones realizadas).

Existen diferentes criterios para clasificar los datos. En función de los métodos utilizados para su recolección, y dependiendo de si se tiene acceso a toda la población objeto de estudio o solo a una parte de ella (muestra), podemos diferenciar entre datos censales y datos obtenidos de encuestas (Nirel & Glickman, 2009). A su vez, dependiendo de cómo se ha seleccionado la muestra en una encuesta podemos distinguir entre métodos probabilísticos y no probabilísticos (Pimienta, 2000). En este documento se exponen las técnicas que se han utilizado para obtener/generar tres conjuntos de datos, atendiendo al criterio de clasificación comentado.

Datos censales

Podemos definir censo como el recuento de datos estadísticos sobre una población específica (Nirel & Glickman, 2009). Teniendo acceso a toda la población se pueden conocer, por tanto, sus características con precisión. Este tipo de estudios suelen ser utilizados por los Gobiernos (nacionales, autonómicos y locales) para conocer la población ubicada en su territorio, atendiendo a una serie de variables o

características sociodemográficas (edad, estado civil, estudios...). Cuando se dispone de esta información de forma recurrente se pueden establecer comparaciones temporales/longitudinales, si bien es cierto que manejar estos conjuntos de datos puede resultar complejo, debido a su gran magnitud. Algunos ejemplos que encontramos en España son el Censo de Población, el Padrón Municipal y el Censo Electoral.

En los sistemas democráticos, los resultados electorales representan la manifestación de la voluntad popular (Biermann & Gupta, 2011). La publicación de estos resultados es fundamental para dar al sistema la legitimidad necesaria. En los últimos años, y gracias sobre todo a los medios de comunicación, es relativamente sencillo consultar/encontrar en la red los resultados de escrutinio hasta el nivel municipal. Esta información, sin embargo, presenta dos importantes restricciones que limitan la realización de análisis a medida. Por una parte, las posibilidades de consulta no cubren niveles de agregación inferior. Por otra parte, los datos no suelen estar disponibles para su descarga, por lo que su tratamiento sistemático, a partir de estas fuentes, requiere del uso de técnicas de *web scraping* (Munzert et al., 2015).

En el caso de los procesos electorales organizados por el (responsabilidad del) gobierno de España (elecciones al Congreso y al Senado, elecciones municipales, elecciones al Parlamento Europeo, elecciones a cabildos o referéndums), los resultados están disponibles para su descarga en la página web del Ministerio del Interior¹. Si bien es cierto que, para la mayoría de las elecciones, el detalle de los datos disponibles en esta web llega incluso a nivel de mesa electoral, el mínimo nivel de agregación en que se cuentan y publican votos en España, el formato en el que se presentan es poco amigable (no estándar), por lo que el investigador debe realizar una serie de adaptaciones para poder trabajar con ellos. Para la mayoría de los ciudadanos y gran parte de los científicos sociales estos archivos constituyen unos jeroglíficos difíciles de resolver. En general, estos agentes no disponen de los conocimientos informáticos necesarios para transformar los archivos: convertir los datos en información valiosa que puedan analizar. Los archivos para descarga se componen de varios ficheros ASCII (con posiciones fijas y estructuras *long-table*) llenos de códigos numéricos que etiquetan toda la información. De hecho, es preciso recurrir a los documentos de metadatos para interpretar y transformar a formatos más intuitivos el contenido de dichos ficheros ASCII (Bonfitto et al., 2021).

Para resolver este problema, hemos descargado todos los microdatos disponibles en la página web del Ministerio y, mediante scripts de R elaborados ad-hoc, hemos generado libros de trabajo (*workbooks*) de datos estructurados. Estos ficheros se han completado añadiendo los niveles de agregación superiores (sección, distrito, municipio, provincia y Comunidad Autónoma), utilizando nuevamente scripts de R. Este método puede resultar no excesivamente complejo, pero cuando se manejan cientos de ficheros, automatizar los procesos es fundamental.

Además de lo anterior, en España, debido a la delegación de ciertas competencias que atribuye el Estado a las Comunidades Autónomas, el tratamiento de datos asociado a las Elecciones Autonómicas presenta una complejidad añadida, que reside en las múltiples fuentes de información existentes. Hay comunidades autónomas que disponen de páginas web estables donde ofrecen los ficheros de resultados (eso sí, cada una con sus formatos y sus peculiaridades). Otras ponen fugazmente los datos a disposición del público, para luego hacerlos desaparecer sin previo aviso, o ir más allá haciendo desaparecer la

¹ <https://infoelectoral.interior.gob.es/>

página web que los alojaba o el organismo encargado. Hay comunidades autónomas que incluso nunca hacen públicos sus resultados electorales detallados y conseguirlos se convierte, a veces, en una carrera de obstáculos, con trabas crecientes a medida que nos alejamos del día de las elecciones, cuando ha habido un cambio de gobierno o por un sentido de propiedad de los datos mal entendido. Averiguar la persona/servicio/organismo a la que hay que hacer la petición y poder conseguir finalmente los datos puede llevar meses, incluso años.

Por todo lo anterior, hemos considerado que resultaría de gran utilidad, tanto para la comunidad científica como para la sociedad en general, y siguiendo con el propósito de generar datos de calidad, que nos permitan ahondar en el conocimiento de la sociedad, crear the Spanish Electoral Archive (SEA) Database, una gran base de datos que reúne de manera abierta, gratuita y homogénea todos los datos recopilados por GIPEyOP durante años, quedando disponibles para cualquier agente a través de la página web del grupo de investigación² y siendo accesible también desde el repositorio Harvard Dataverse³. Los detalles de esta base de datos, que contiene en el momento de redacción del presente documento casi 1000 ficheros, alcanzando algunos de ellos 60000 observaciones y más de 100 variables, están descritos en profundidad en la publicación *Spanish electoral archive. SEA database* (ver Anexo I), la primera de las publicaciones que conforman esta Tesis doctoral.

Datos obtenidos de encuestas (métodos probabilísticos)

Cuando el investigador no tiene acceso a toda la población objeto de estudio, o le resulta muy costoso/complejo hacerlo, puede utilizar solo una parte de ella, seleccionando una muestra que represente, en la mayor medida posible, a dicha población (Gideon, 2012). En este sentido, las encuestas sociodemográficas se han convertido en una fuente de información muy valiosa, tanto para agentes públicos y privados como para investigadores. La utilidad de esta fuente de información a la hora de analizar las características de la población total y/o de subpoblaciones, como la económicamente activa, es incuestionable (Oliveira & García, 1987). En la mayoría de los países desarrollados existen registros periódicos de encuestas recopiladas por institutos/centros de investigación nacionales, las cuales permiten obtener información sobre, por ejemplo, ocupación/desocupación y/o ingresos/gastos, permitiendo incluso captar información más completa que la obtenida de los censos de población (Deaton, 2000). Además de lo anterior, las posibilidades que ofrecen los datos de encuesta son múltiples, permitiendo, entre otras, la construcción de modelos estadísticos que buscan detectar la importancia relativa de los múltiples factores condicionantes de las variables demográficas (Lumley & Scott, 2017).

¿Cómo ha cambiado la sociedad española en las últimas décadas? En España, disponemos de un banco de datos que puede permitir abordar esta cuestión desde una perspectiva longitudinal, observando la evolución del comportamiento de nuestros conciudadanos a lo largo de los últimos 40-50 años. Un repositorio que alberga un gran volumen de datos, y que, por tanto, puede resultar de enorme interés para la sociedad, además de poseer una considerable relevancia científica. No obstante, este banco de datos se considera sustancialmente inexplorado, posiblemente por la dificultad que supone manejar tal

² <http://sea.uv.es/gipevop/sea.html>

³ <https://dataverse.harvard.edu/dataverse/SEA>

volumen de datos. Nos referimos al banco de datos que el Centro de Investigaciones Sociológicas (CIS) pone a disposición de la comunidad científica y de la sociedad en general en su página web oficial⁴.

Las encuestas realizadas por el CIS son un referente de los estudios de opinión en España (Pavía & García-Cárceles, 2012) recogiendo de manera periódica y continuada información relacionada con, entre otras, i) variables sociodemográficas básicas; ii) valoración de la situación económica; iii) variables políticas; iv) actitudes personales; v) nivel de estudios; vi) estado civil y situación de convivencia; vii) variables laborales y socio-económicas; además de un conjunto de variables específicas correspondientes a un tema de actualidad, en el momento de realización del estudio. Estas encuestas generan un elevado volumen de datos: más de cuatro millones de respuestas recolectadas en los más de 2200 estudios disponibles actualmente (el 80% con un tamaño muestral superior a 1000, existiendo estudios que alcanzan $N \approx 30000$), con una media de 150 variables observadas en cada estudio. Y esta cifra no para de crecer, puesto que, además de que cada año se recopila nueva información, el CIS está completando la base de datos hacia atrás en el tiempo, incorporando de manera gradual microdatos de encuestas realizadas en los años ochenta y noventa del siglo pasado. Por todo lo anterior, el banco de datos del CIS constituye el repositorio de encuestas más completo y amplio de España y uno de los más valiosos de cuantos se pueden encontrar en las democracias occidentales (Pavía & Aybar, 2018).

Manejar un conjunto de datos de tales dimensiones no es sencillo. Es preciso trabajar de forma sistemática y automatizando procesos. En el caso concreto de la información que ofrece el CIS, hemos considerado apropiado elaborar un mapa temático (empleando el software especializado R), utilizando técnicas de *text mining* (Weiss et al., 2005) para extraer las palabras clave (Onan et al., 2016; Rose et al., 2010) y de procesamiento de lenguaje natural para establecer significados (Sun et al., 2017), con un doble objetivo. Primero, identificar cada pregunta dentro de cada estudio y, segundo, asociar preguntas y temáticas entre estudios. El resultado de este proceso ha sido un diccionario que facilita moverse entre estudios y preguntas, y así abordar cuestiones a medida. Esta herramienta, fruto de un enorme trabajo de identificación de temáticas y variables, abre una gran ventana de posibilidades de análisis, y reduce considerablemente los costes de búsqueda y procesamiento de datos. Entre otras cuestiones, facilita abordar la resolución de problemas como la elaboración de tablas ideológicas cohorte-generacionales condicionadas a diferentes variables, las cuales permitirían evaluar escalas de valores y observar cómo han ido cambiando según la ideología de los ciudadanos a lo largo de los últimos años.

Disponer de esta valiosa herramienta nos ha permitido elaborar dos papers más (uno de reciente publicación y otro en estado de revisión), circunstancia que nos ayuda a visibilizar su enorme utilidad. En la primera de estas dos publicaciones, que tiene por título *Expanding the New Roles of the Military. The Case of Spain's Military Emergency Unit* (ver Anexo II) analizamos la posible relación entre los nuevos roles que están desempeñando las Fuerzas Armadas en España y la percepción que tiene la sociedad española de sus militares. Este estudio se ha fundamentado en datos de 12 encuestas del CIS, realizadas entre 1997 y 2017. Es muy probable que, de no haber tenido el mapa temático (diccionario), este estudio nunca se hubiera realizado, puesto que detectar en qué encuesta se plantean determinadas cuestiones, dentro de un banco de datos que aglutina más de 2200 estudios, es prácticamente imposible. Y no solo eso. Puesto que en cada estudio existen unas preguntas que son comunes y otras que no lo

⁴ https://www.cis.es/cis/opencms/ES/2_bancodatos/

son, crear un *dataset* que aglutine la información de los 12 estudios no es inocuo. Conocer qué preguntas (variables) se repiten en las encuestas seleccionadas resulta, por tanto, de gran utilidad.

La segunda publicación que hemos realizado utilizando nuestro diccionario es *Escalas de Likert, auto-ubicación ideológica y encuestas del CIS* (ver Anexo III), documento que está en revisión en el momento de redacción de esta Tesis doctoral. Los más de 2200 estudios que ofrece el CIS en su banco de datos se pueden clasificar en cinco grandes bloques: i) Barómetros; ii) Índices de Confianza del Consumidor; iii) Encuestas electorales; iv) Estudios sobre temas políticos; y v) Otros estudios. El método para realizar la entrevista varía dependiendo del tipo de estudio (presencial, telefónica o auto-administrada por internet). Esto permite establecer comparaciones y detectar posibles diferencias en ciertas variables, dependiendo del tipo de estudio y/o el modo de administración.

En el 90% de los estudios, el CIS pregunta sobre la autoubicación ideológica, pero a lo largo de los años lo ha hecho de forma diferente, utilizando hasta 28 formas distintas de medición (escalas Likert), por lo que analizar la evolución de esta variable, o averiguar si influye la temática y/o el tipo del estudio, no es trivial. Concretamente, en este trabajo hemos tenido que redefinir la variable IDEOLOGIA en 28 nuevas variables, teniendo en cuenta las opciones de respuesta que se ofrecían en cada estudio y la forma de realizar/plantear la pregunta. Nuevamente, la herramienta que hemos creado nos ha facilitado sobremanera el trabajo, convirtiéndose en un recurso fundamental para futuros proyectos.

Datos obtenidos de encuestas (métodos no probabilísticos)

Si bien es cierto que las encuestas online se realizan con cierta asiduidad desde los años 90 del siglo pasado, su uso se ha extendido en España en la última década, cuando el acceso a internet por parte de la población se ha generalizado, convirtiéndose el correo electrónico, la mensajería instantánea y el uso de redes sociales en los canales de comunicación más utilizados (Arroyo Menéndez & Finkel, 2019). Este hecho, sumado al mayor coste que implica la realización de encuestas personales/telefónicas, o la imposibilidad de realizarlas en determinadas circunstancias, está motivando un incremento sustancial del uso de métodos no probabilísticos en encuestas. Así, no es de extrañar que estemos asistiendo a la aparición de nuevas experiencias, inicialmente en el mundo académico y mediante el uso de paneles, ya ampliamente aceptadas en el sector, que, aprovechando las posibilidades de Internet, tratan de demostrar que también es posible generar inferencias de calidad con muestras sesgadas, empleando, por ejemplo, mecanismos en los que la población potencialmente encuestada no es seleccionada por el encuestador, sino que los encuestados se auto seleccionan (Pimienta, 2000).

GIPEyOP ha realizado varios estudios de esta naturaleza, los dos más recientes en 2020, en plena pandemia de la COVID-19 (durante el confinamiento domiciliario y tras finalizar un periodo estival sin apenas restricciones de movilidad). Una vez elaborado, el cuestionario autoadministrado se envía, mediante correo electrónico, mensajería instantánea y redes sociales, a los colaboradores del grupo de investigación, procurando que estos, a su vez, se lo faciliten a sus propios contactos, y así sucesivamente. El atractivo de este método no probabilístico, denominado bola de nieve, radica en el número de personas (tamaño muestral) que participa en la encuesta, circunstancia que compensa los sesgos de este método (tasas de no respuesta elevadas y baja representatividad de la población, fundamentalmente).

Aunque este tipo de estudios tiene carencias, resulta de utilidad para investigadores en áreas muy diversas. En el caso concreto de las dos encuestas mencionadas anteriormente, los datos obtenidos

pueden ser utilizados, entre otros, i) por economistas, sociólogos y politólogos, para evaluar las teorías de género sobre el comportamiento dentro del hogar en un escenario de desdibujamiento de los roles de género tradicionales (estricto confinamiento domiciliario); ii) por economistas, para evaluar cómo los posibles cambios de hábitos impactan/han impactado en la economía; iii) por politólogos, para analizar si las políticas implementadas por los gobernantes impactaron/han impactado en las preferencias de voto de los ciudadanos; iv) por psicólogos, para evaluar cómo influye el miedo en el comportamiento de los españoles. Además de lo anterior, estos estudios visibilizan la utilidad de las encuestas realizadas con muestras no aleatorias, de las que se puede obtener información muy valiosa.

Tanto los cuestionarios utilizados como los resultados obtenidos (microdatos) de los dos estudios realizados se han puesto a disposición de la comunidad científica como material suplementario de las publicaciones *Dataset of the COVID-19 lockdown survey conducted by GIPEyOP in Spain* y *Dataset of the COVID-19 post-lockdown survey conducted by GIPEyOP in Spain* (ver Anexos IV y V). Estos dos papers forman parte del compendio de publicaciones que conforman la presente Tesis doctoral.

Además de lo anteriormente expuesto, y a pesar de que los encuestados que han participado en ambas encuestas no tienen por qué ser los mismos, el planteamiento de las preguntas formuladas permite analizar el comportamiento de la sociedad en distintos momentos del tiempo, observando un posible cambio de hábitos. Esta circunstancia nos ha llevado a observar si el miedo es uno de los posibles factores que ha condicionado el comportamiento de la población, centrándonos en cómo y con qué frecuencia usan (usaban) los servicios de restauración (bares y restaurantes, fundamentalmente). Los análisis realizados y las conclusiones obtenidas se expusieron en el *Simposio sobre los impactos sociales del COVID-19*, organizado por la Federación Española de Sociología, y se presentan en el paper *COVID-19 and Changes in Social Habits. Restaurant Terraces, a Booming Space in Cities. The Case of Madrid* (ver Anexo VI), el cual también forma parte de esta Tesis doctoral.

Conclusiones y futuros trabajos

La revolución digital permite recolectar una cantidad ingente de datos, muy útil para el progreso científico y social. No obstante, no basta con tener acceso a ese gran volumen de datos. Los datos deben de ser de calidad para que las conclusiones que se extraigan de ellos sea relevante, permitiendo así una toma de decisiones adecuada y generar conocimiento. En este sentido, consideramos fundamental implementar procedimientos y utilizar herramientas que permitan transformar los datos brutos en datos de calidad, favoreciendo la labor del investigador, realizando su función de la manera más eficiente posible.

Si bien es cierto que las autoridades, tanto nacionales como supranacionales, están trabajando en homogeneizar la forma en que se presentan los datos abiertos en los repositorios web públicos, estableciendo unos mínimos de calidad y definiendo unas normas básicas de actuación, las medidas propuestas no dejan de ser recomendaciones, por lo que este proceso de unificación de criterios no se está llevando a cabo al ritmo que cabría esperar. Por ello, facilitar al investigador herramientas que le permitan adaptar los datos a sus necesidades es crucial.

Las herramientas y métodos propuestos en este documento facilitan el trabajo científico, reduciendo los esfuerzos que el investigador tiene (tenía) que dedicar en la fase de preparación de los datos,

pudiéndose centrar así en las labores propias de su campo/área de conocimiento. Construir grandes bases de datos a partir de conjuntos de microdatos aislados o crear un mapa temático que contiene todas las variables generadas por miles de encuestas han sido algunos de los outputs que hemos obtenido. No obstante, nuestro trabajo no acaba aquí, de hecho, recién acaba de empezar. A continuación, se exponen algunas ideas que estamos desarrollando en estos momentos, así como algunos proyectos que tenemos intención de llevar a cabo a corto plazo:

- **Control de calidad.** Como se ha expuesto en el presente documento, partiendo de los microdatos originales, hemos creado una base de datos agrupando observaciones en función de una determinada variable (base de datos electoral con diferentes niveles de agregación). Las cifras obtenidas en esta nueva base de datos, construida a partir de los resultados electorales a nivel de mesa, pueden ser comparadas con las cifras publicadas en los diarios y boletines oficiales correspondientes (normalmente a nivel provincia), circunstancia que permitiría comprobar la existencia de errores y, en tal caso, dilucidar de qué tipo son y si son subsanables, así como proponer métodos para su corrección. Es más, utilizando técnicas de auditoría de datos y de aprendizaje automático podremos identificar y cuantificar el tamaño y alcance de los errores que no se corrigen pero que son evidentes en un análisis estadístico.
- **Añadir nuevas variables a nuestras bases de electorales.** Existen multitud de variables socioeconómicas, disponibles a nivel de sección censal, que podrían ser vinculadas con los resultados electorales disponibles. Incorporar nuevas variables aportaría un gran valor a las bases de datos generadas y facilitaría sobremanera la realización de nuevos trabajos de investigación.
- **Incorporar información espacial.** En el momento de la realización de esta Tesis doctoral, menos de la mitad de los datos disponibles en abierto disponen de información geográfica (Abella et al., 2021). Conocer qué ocurre en qué lugar resulta de gran interés para diversos agentes sociales y económicos, siendo la segunda categoría de datos más reutilizada y consultada por las empresas en la UE, según el informe Re-using Open Data (European Data Portal, 2020). Por ello, completar nuestras bases de datos con información de este tipo será una de las primeras tareas a abordar, siguiendo con el trabajo iniciado en la base de datos SEA, en la que ya ofrecemos un *shapefile*.
- **Ofrecer el mapa temático (diccionario) del CIS a la comunidad científica.** Tal y como se ha comentado en este documento, una de las herramientas más valiosas que hemos creado ha sido el diccionario del CIS. Ponerlo a disposición de la comunidad científica y de la sociedad en general es una de nuestras prioridades. No obstante, nuestra intención es hacerlo cuando hayamos depurado algunos aspectos todavía pendientes de tratamiento y revisión, siendo conscientes de que nunca va a estar acabado, fundamentalmente por dos motivos: i) el CIS incorpora nuevos estudios cada pocas semanas; y ii) nuevas investigaciones permiten detectar/generar nuevas variables, y así ha ocurrido en la realización del trabajo *Escalas de Likert, auto-ubicación ideológica y encuestas del CIS*. Además de lo anterior, revisar y actualizar el diccionario permite (y ha permitido) detectar (posibles) errores en los microdatos ofrecidos por el CIS (variables mal codificadas o datos faltantes, por ejemplo), información que le es trasladada de inmediato a los responsables del banco de datos para su corrección.
- **Construir tablas ideológicas cohorte-generacionales.** Una de las cualidades de los datos del CIS es que los estudios se ofrecen con regularidad, planteando diversas cuestiones de forma

recurrente, lo que permite realizar estudios longitudinales con un volumen de datos considerable. Observar el comportamiento de determinadas variables (como la ideología política) con el paso del tiempo y para sujetos de diferentes edades puede ser de gran interés, siendo uno de nuestros actuales proyectos la construcción de tablas ideológicas cohorte-generacionales. Para la construcción de estas tablas estamos utilizando técnicas de doubles ventanas móviles (edad-año), que representan una extensión metodológica de la técnica de ventana móvil utilizada en otros contextos. Generar estas tablas nos permitirá realizar análisis sobre la evolución de distintos marcadores, por ejemplo, la modernización que ha habido en las últimas décadas y el impacto generado, sobre todo, en las personas mayores. Los avances realizados hasta el momento se han expuesto en varios congresos científicos, tanto nacionales como internacionales.

- **Realizar un análisis más completo de la percepción que tiene la sociedad española de sus Fuerzas Armadas.** En la publicación *Expanding the New Roles of the Military. The Case of Spain's Military Emergency Unit*, presentada en este documento, y que forma parte del compendio de artículos que conforma la presente Tesis doctoral, se observa que, a partir de 2005, año de creación de la UME, la valoración de la institución militar por parte de los españoles mejora. Si bien es cierto que no podemos demostrar causalidad entre ambos hechos, los datos con los que hemos trabajado sí son compatibles con esa hipótesis. Por tanto, sería interesante, por un lado, generar/proponer modelos estadísticos que permitan comprobar qué variables, y en qué medida, influyen en la valoración que hacen los españoles de la institución militar y, por otro lado, obtener/recolectar nuevos datos, que complementen a los ya disponibles, a partir de encuesta(s) elaborada(s) por GIPEyOP. Esto permitiría, por ejemplo, detectar si la valoración que hacen los españoles de las Fuerzas Armadas, en su conjunto, es distinta a la percepción que tienen de la Unidad Militar de Emergencias.

Bibliografía

- Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2018). Indicadores de calidad de datos abiertos: el caso del portal de datos abiertos de Barcelona. *Profesional de la Informacion*, 27(2), 375–382. <https://doi.org/10.3145/epi.2018.mar.16>
- Abella, A., Ortiz-de-Urbina-Criado, M., De-Pablos-Heredero, C., & García Luna, D. (2021). *La Reutilización de Datos Abiertos III*. https://www.meloda.org/wp-content/uploads/2022/03/Datos_abiertos_en_spain_III_2021.pdf
- Arroyo Menéndez, M., & Finkel, L. (2019). Encuestas por Internet y nuevos procedimientos muestrales. *Panorama Social*, 30, 41–53.
- Biermann, F., & Gupta, A. (2011). Accountability and legitimacy in earth system governance: A research framework. *Ecological Economics*, 70(11), 1856–1864. <https://doi.org/10.1016/j.ecolecon.2011.04.008>
- BOE. (2013). Resolución de 19 de febrero de 2013, de la Secretaría de Estado de Administraciones Públicas, por la que se aprueba la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información. *Boletín Oficial Del Estado*, 54, 17045–17071.
- Bonfitto, S., Casiraghi, E., & Mesiti, M. (2021). Table understanding approaches for extracting

- knowledge from heterogeneous tables. *WIREs Data Mining and Knowledge Discovery*, 11(4). <https://doi.org/10.1002/widm.1407>
- Curto-Rodríguez, R. (2021). Análisis multidimensional de los portales de datos abiertos autonómicos españoles. *Revista Española de Documentación Científica*, 44(1). <https://doi.org/10.3989/redc.2021.1.1745>
- Deaton, A. (2000). *The Analysis of Household Surveys* (3rd ed.). The Johns Hopkins University Press.
- European Data Portal. (2020). *Re-using Open Data*. <https://doi.org/10.2830/876679>
- Gideon, L. (Ed.). (2012). *Handbook of Survey Methodology for the Social Sciences*. Springer New York. <https://doi.org/10.1007/978-1-4614-3876-2>
- Guillén, M. A., López Ayuso, B., Paniagua, E., & Cadenas, J. M. (2015). Una revisión de la Cadena Datos-Información-Conocimiento desde el Pragmatismo de Peirce. *Documentación de Las Ciencias de La Información*, 38, 153–177. https://doi.org/10.5209/rev_dcin.2015.v38.50814
- Lumley, T., & Scott, A. (2017). Fitting Regression Models to Survey Data. *Statistical Science*, 32(2). <https://doi.org/10.1214/16-STS605>
- Munzert, S., Rubba, C., Meibner, P., & Nyhuis, D. (2015). *Automated data collection with R: a practical guide to web scraping and text mining*. Wiley.
- Nirel, R., & Glickman, H. (2009). Sample Surveys and Censuses. In C. R. Rao (Ed.), *Handbook of Statistics* (pp. 539–565). Elsevier. [https://doi.org/10.1016/S0169-7161\(08\)00021-7](https://doi.org/10.1016/S0169-7161(08)00021-7)
- Oliveira, O., & García, B. (1987). Encuestas, ¿hasta dónde? *Revista Mexicana de Sociología*, 49(1), 335–351.
- Onan, A., Korukoğlu, S., & Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. *Expert Systems with Applications*, 57, 232–247. <https://doi.org/10.1016/j.eswa.2016.03.045>
- Pavía, J. M., & Aybar, C. (2018). Field rules and bias in random surveys with quota samples. An assessment of CIS surveys. *SORT*, 42(2), 183–206. <https://doi.org/10.2436/20.8080.02.74>
- Pavía, J. M., & García-Cárceles, B. (2012). Una Aproximación Empírica al Error de Diseño Muestral en las Encuestas Electorales del CIS. *Metodología de Encuestas*, 14, 45–62.
- Peset, F., Aleixandre-Benavent, R., Blasco-Gil, Y., & Ferrer-Sapena, A. (2017). Datos abiertos de investigación. Camino recorrido y cuestiones pendientes. *Anales de Documentación*, 20(1), 1–12. <https://doi.org/10.6018/analesdoc.20.1.272101>
- Pimienta, R. (2000). Encuestas probabilísticas vs. no probabilísticas. *Política y Cultura*, 13, 263–273.
- Poisot, T., Bruneau, A., Gonzalez, A., Gravel, D., & Peres-Neto, P. (2019). Ecological Data Should Not Be So Hard to Find and Reuse. *Trends in Ecology & Evolution*, 34(6), 494–496. <https://doi.org/10.1016/j.tree.2019.04.005>
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia Artificial*, 10(29), 11–18.
- Rivas-Rebaque, B., Gétrudix-Barrio, F., & De Britto, J. C. D. C. (2019). La percepción del docente universitario ante el uso y valor de los datos abiertos. *Educacion XX1*, 22(2), 141–163. <https://doi.org/10.5944/educxx1.21317>

- Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. *Text Mining: Applications and Theory, 1*, 1–20.
- Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion, 36*, 10–25. <https://doi.org/10.1016/j.inffus.2016.10.004>
- Verna, N., & Gupta, M. P. (2012). Open Government Data: More than Eighty Formats. In *Towards E-governance in the Cloud* (pp. 207–216). https://www.researchgate.net/profile/Devika-Palanisamy/publication/331287764_E-Governance/links/5c707c1f299bf1268d1e095b/E-Governance.pdf#page=214
- Weiss, S. M., Indurkha, N., Zhang, T., & Damerou, F. J. (2005). *Text mining: predictive methods for analyzing unstructured information*. Springer Science+Business Media.

Anexo I

Spanish electoral archive. SEA database

Pérez, V., Aybar, C. & Pavía, J. M. (2021). *Scientific Data*, 8(1), 193.

<https://doi.org/10.1038/s41597-021-00975-y>



OPEN

Spanish electoral archive. SEA database

DATA DESCRIPTOR

Virgilio Pérez¹, Cristina Aybar¹ & Jose M. Pavía²✉

This paper introduces the SEA database (acronym for Spanish Electoral Archive). SEA brings together the most complete public repository available to date on Spanish election outcomes. SEA holds all the results recorded from the electoral processes of General (1979–2019), Regional (1989–2021), Local (1979–2019) and European Parliamentary (1987–2019) elections held in Spain since the restoration of democracy in the late 70s, in addition to other data sets with electoral content. The data are offered for free and is presented in a homogeneous and friendly format. Most of the databases are available for download with data from various electoral levels, including from the ballot box level. This paper details how the information is organized, what the main variables are on offer for each election, which processes were applied to the data for their homogenization, and discusses future areas of work. This data has many applications, for example, as inputs in election prediction models and in ecological inference algorithms, to study determinants of turnout or voting, or for defining marketing strategies.

Background & Summary

In an open, modern, transparent society, access to detailed and reliable information is essential. Basic data should not only be accessible, but should also be available in user-friendly formats. In democratic systems, electoral results represent the manifestation of the popular will. Publication of these results is essential to give the system the necessary legitimacy. In the case of Spain, despite the so-called Law of Transparency¹, gaining access to detailed electoral data can sometimes be convoluted and, even once access is gained, the format of the data often needs to be edited and homogenized before being usable.

Election data is a valuable source of information for answering a multitude of research questions, profiling citizens, and implementing business models. Electoral behaviour varies depending on the region, the province, the municipality, the neighbourhood, the district, the census section or the ballot box. Having detailed results for small areas (for example, at the municipal, section or polling station level) is relevant for researchers, analysts and professionals. For example, the information is useful for the definition of electoral marketing strategies², for geographic analysis³, to generate estimates with data from pre-electoral surveys using models of small areas⁴, to improve the predictions of exit polls^{5–7} and quick counts⁸, and for use in spatial or spatio-temporal electoral models⁹. The previous examples are not exhaustive; the data can also be used, among other things, as inputs in ecological inference algorithms^{10–12}, in electoral night prediction models^{13,14}, to study determinants of participation or voting^{15,16}, or for redrawing of the electoral spaces^{17,18}.

In the case of electoral processes organized by, and the responsibility of, the government of Spain (elections to Congress and the Senate, municipal elections, European Parliamentary elections, elections to councils or referendums), the results are available for direct download in the “Área de Descargas” from the website of the Spanish Ministry of the Interior. For most elections, data available on this website can be as detailed as at the ballot box level, the minimum level of aggregation in which votes are counted and published in Spain.

For most citizens and a number of social scientists, however, these files are nothing more than strange hieroglyphics that are difficult to read. In general, these agents do not have the necessary computer skills to transform files, i.e., turn data into valuable information that they can analyse. The files for download are composed of several ASCII files (with fixed positions and long-table structures) filled with numerical codes that label the information. Metadata documents are necessary in order to interpret and transform the content of ASCII files into more intuitive formats.

In the case of regional elections, the access to the data is even more problematic. There are regions (autonomous communities) that have active web pages in which they offer files of the results (each file with its own

¹GIPEyOP, Área de Métodos Cuantitativos, Departamento de Economía Aplicada, Facultat d’Economía, Universitat de Valencia, Valencia, Spain. ²GIPEyOP, UMICCS, Área de Métodos Cuantitativos, Departamento de Economía Aplicada, Facultat d’Economía, Universitat de Valencia, Valencia, Spain. ✉e-mail: pavia@uv.es

format and characteristics). Others make the data available briefly to the public, then make it disappear without prior notice, or go one step further by making the page that hosted it or the agency in charge disappear (see examples in Online-only Table 1). There are autonomous communities that never even make their detailed electoral results public, and getting access to this information is, at times, far from easy. It becomes even harder the further removed the date is from the election date, or when there has been a change of government or as a consequence of a misunderstood sense of ownership of data. Finding out who to contact to make the request, be it person/service/body, is difficult but, once done, actually getting the data can take months, even years.

Undoubtedly, many agents would find it useful to have a centralized location where they can access, in a homogeneous and friendly format, all the electoral data available for Spain. The research group on Electoral Processes and Public Opinion of the University of Valencia (GIPEyOP) has been undertaking this task for a number of years (sometimes driven by our own needs), collecting and homogenizing the information associated with the main electoral processes held in Spain. All the data collected by GIPEyOP <http://gipeyop.uv.es/> are available in an open, free and homogeneous format in the SEA database (acronym for: Spanish Electoral Archive) and are accessible through the web pages <http://sea.uv.es/gipeyop/sea.html> and <https://dataverse.harvard.edu/dataverse/SEA>. This article describes the structure and variables that can be found in the different files that make up SEA, and reports on the enrichment and improvement processes that GIPEyOP is undertaking to expand the quality and scope of the data.

Methods

The SEA Database has been developed from processing and standardizing official (provisional and proclaimed) outcomes available from different sources and in different formats. The primary data sources and the processing performed are detailed below, grouped by type of election.

General elections. In the case of the General Elections, SEA offers for download all the results corresponding to the elections to the Congress of Deputies held in Spain since the restoration of democracy. This represents a total of 15 elections at the time of writing this article: from the elections held in June 1977 to the last of November 2019. The primary source of information for the data of these elections is the Ministry of the Interior of Spain in the block “*Extracción de datos*” from the “*Área de Descargas*” section: <http://www.infoelectoral.mir.es/infoelectoral/min/areaDescarga.html>.

In each electoral process, the Spanish Ministry of the Interior offers, for download, data at the maximum level of detail available (usually at the ballot box level). The download is offered in the form of a compressed file (zip) containing 12 files: 10 ASCII files, with a DAT extension, 1 Microsoft Word file, with a doc extension, and a final rich text file, with an rft extension. The 10 DAT files have a fixed length, text format and a constant structure, which facilitates the programming load for their reading and subsequent compilation. The doc and rft files are metadata files that help with understanding the content and structure of the information in DAT files.

Specifically, there are four DAT files that must be processed to get some of the variables of interest listed in the next section: the 03xxymm.DAT file, from which information on the candidatures (codes, acronyms and party names) is obtained; the 05xxymm.DAT file, which offers us information on the spatial identifiers of the voting units (codes and names); the 09xxymm.DAT file, which contains, among other variables, census information and information on blank and null votes; and the file 10xxymm.DAT that collects the votes per candidacy. In these file names ‘xx’ tells us about the type of electoral process (for example, when xx = 02, this signifies elections to the Congress of Deputies, xx = 04 Municipal elections and xx = 05 elections to the European Parliament), and ‘ymm’ the year and the month when the election was held.

All the data processing to transform the unfriendly data formats and metadata of the Ministry of the Interior into structured data workbooks has been done using ad-hoc scripts in R¹⁹, executed through RStudio²⁰. Once all the information contained in files 03, 05, 09 and 10 has been extracted, the information for the higher aggregation scopes is generated (again with ad-hoc R scripts).

Regional elections. The data processing associated with the Regional (Autonomous) Elections presents a further complexity to that of other elections, a complexity that resides in its multiple sources of information. Management of the electoral processes in each autonomous community is self-governing and does not depend on the Ministry of the Interior. The difficulty is that not all autonomous governments offer the results on public platforms. In some cases, it is the Statistical Institutes of the corresponding autonomous community which offer the information, in other cases this information is not available. The procedure to follow when the information is not public, or is not easily located, is to contact the person responsible for the electoral processes of the corresponding administration. Finding the right contact, however, can be a lengthy, and not always successful, process.

Sometimes the results are obtained after being referred to the company that was hired to manage the entire counting process. Being dependent on the company contracted by each autonomous government can be problematic because the company chosen may change following an election win or after changes of government. This means there is a possibility that the links that we offer in our processed files as the primary source of data may become obsolete. Likewise, there have been autonomous governments that have provided us with information directly from a representative of their administration, so in these cases we cannot offer the link to the original raw data. Online-only Table 1 offers information on the links of the autonomous communities and their current status.

The content of Online-only Table 1 is a clear reflection of the difficulties involved in merely obtaining electoral results at the autonomous community level and its heterogeneous level of transparency in relation to electoral data. The information presented in Online-only Table 1 gives an idea of the haphazard approach to offering access to data and the problems that have arisen in the process of collecting electoral information through regional practices, difficulties which have made the process of creating SEA all the more challenging, but necessary.

Local elections. The primary source of information of the results of municipal elections is the same as that of the European and General elections: the Ministry of the Interior. Municipal elections are held every 4 years. To date, these elections have been held in the following years: 1979, 1983, 1987, 1991, 1995, 1999, 2003, 2007, 2011, 2015 and 2019. SEA (the Ministry of the Interior) has completed information (that is, from ballot box level) since the 1987 elections. For the first two elections, 1979 and 1983, the results are only available since municipal (local) level.

The advantage of having the Ministry of the Interior as a source of information is that it always presents the same structure, the same text files, and this facilitates data processing, which is why similar procedures to those described above were used. The difficulty, in this case, lies in the fact that in each municipality different candidacies are presented, and extra programming efforts are required to resolve this issue.

European Parliamentary elections. The electoral years corresponding to European Parliamentary elections are: 1987, 1989, 1994, 1999, 2004, 2009, 2014 and 2019. For all of these there is raw data per ballot box and the source of information is, as in the other national electoral processes, the Ministry of the Interior. The data processing is similar to those described for the general elections, although with some particular features.

In these elections, despite the fact that Spain represents a single constituency, a large number of different candidacies are usually presented, which makes it difficult to achieve representation. The electoral law²¹, however, allows different candidates to form coalitions (groupings) that must be considered together when distributing MEPs. In this way, the votes received by candidacies that use different acronyms and party names in different territories are added together before applying the electoral rules that transform the votes into representatives²². This idiosyncrasy of these elections has made it expedient to build two files for each election: a file with disaggregated voting data for each of the candidacies presented, and another, more reduced and manageable, which facilitates the processing and interpretation of the data, obtained by merging the parties that are part of the same coalition when it comes to distributing MEPs, although they are presented under different names in different territories.

Other sources of data. In addition to purely electoral results, SEA also collates more heterogeneous information in a dataset called “Spanish Elections-Others” (hereinafter “Others”), which, as it grows in the coming years, may lead to new data sets that will be need to be properly documented and communicated. “Others” includes examples of tailor-made compilations, the result of linking various electoral files or linking an electoral file with other sources of information, such as: cartographic files (shape), which contain the geographical limits of the spatial divisions, or population register data files, both available from the scope of the census sections.

It should be noted that correspondence between census sections of the cartographic, electoral and register files, of the same year, do not always coincide^{23,24} and, moreover, they vary over time²⁵, so any combination of these sources requires previous processing. Details of the processing should be offered as additional information so that anyone using such files is aware of the limitations, and the hypotheses, of the information contained in them. At the time of writing, “Others” only offers a few examples of these uses. These are examples that are characterized by presenting a direct correspondence between sections, for which the simplest version of the proposals made by Pavia-Miralles¹⁴ has been used to establish correspondences (matching sections). The files available in this section will gradually increase as we automate the processes of matching and, above all, documentation.

The advantage of combining information sources is that it offers more analysis possibilities. For example, based on the combination offered in SEA that includes the results of general elections and the population register of 1 January, 2019, an analyst could apply ecological inference techniques²⁶ using covariates, better define the sampling frames in surveys elections²⁷, or study the contextual effects that the number and composition of immigration in each census section have on voting²⁸.

Data Records

The SEA Database is available through the Harvard dataverse repository <https://dataverse.harvard.edu/dataverse/SEA>. It contains 974 files in Microsoft Excel format, as well as some compressed (rar) files. Currently, SEA is structured into 4 thematic datasets, plus a fifth additional data set. The names of the first four data sets identify the typology of elections reflected in the data: “Spanish General Elections”²⁹, “Spanish Regional Elections”³⁰, “Spanish Local Elections”³¹ and “Spanish European Elections”³². The fifth dataset, “Spanish Elections - Others”³³, is more heterogeneous. In the data sets of regional elections and municipal elections, a folder structure has also been used to facilitate access to the files. The autonomous community (regional) elections data set is made up of nineteen folders, one for each of the seventeen autonomous communities (CCAA) plus two additional ones for the autonomous cities of Ceuta and Melilla. Each of the folders is identified with the name of the corresponding region (city). Similarly, the municipal elections data set is (currently) composed of eleven folders, one for each of the years in which elections of this type have been held (every four years, from 1979 to 2019). The use of *Tree View* is recommended to navigate these data sets, as each folder contains dozens of files.

The files for download are generally Microsoft Excel workbooks, with a couple of exceptions: the compressed files available in the municipal elections data set, which contain the Excel workbooks of the voting results corresponding to local elections in different areas (more details later), and some of the files for download to the “Others” data set. In this last data set, all the files corresponding to a specific electoral cartography (including the shape files that define the polygonal lines associated with the census sections and the associated dbf database of electoral outcomes) are offered compressed into a single downloadable file. The description that follows mainly focuses on the content that can be found in the Excel workbooks available in the four main data sets. These workbooks contain the official and provisional election results (recorded during the count), and constitute the main content of SEA.

Overview of the data files. Each Microsoft Excel workbook contains detailed results of a single election and can contain up to ten spreadsheets, with titles that indicate their contents. An initial spreadsheet (PREVIEW)

Siglas	Denominación	Andalucía	Andalucía	Andalucía	Andalucía	Andalucía	Andalucía	Andalucía	Andalucía	Aragón	Aragón	Aragón	Asturias
		01 Almería	01 Cádiz	01 Córdoba	01 Granada	01 Huelva	01 Jaén	01 Málaga	01 Sevilla	02 Huesca	02 Teruel	02 Zaragoza	03 Asturias
ADC	Agrupación De Electores Carlistas												
ADEC	Asociación De Electores De Ceuta												
AEICYU	Agrupación Electoral Indep. Del Campo Y La Ciudad												
AET	Agrupación Electoral De Trabajadores					AET			AET		AET	AET	
AFN	Alianza Foral Navarra												
AN18	Alianza Nacional 18 De Julio	AN18		AN18					AN18		AN18	AN18	AN18
ANEPA.CP	Asoc.Nai. Para El Estudio De Problemas Actuales		ANEPA.CP					ANEPA.CP					
ANV	Acción Nacionalista Vasca												
AP	Federación De Partidos De Alianza Popular	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP	AP
ASA	Agrup.Elec. Y Part.Indep. Acción Social Agraria					ASA							
ASDCI	Alianza Socialista Democrática	ASDCI		ASDCI	ASDCI	ASDCI	ASDCI	ASDCI	ASDCI				ASDCI
BAI	Bloque Andaluz De Izquierda							BAI					
BNPG	Bloque Nacional Popular Gallego												
CAI	Candidatura Andaluza De Izquierda				CAI				CAI				
CAIC	Candidatura Aragonesa Independiente De Centro											CAIC	
CCIA	Centro Izquierda De Albacete												
CIPYE	Candidatos Indep. De La Pequeña Y Mediana Empresa							CIPYE					
CJA	Asociación Círculos Jose Antonio						CJA					CJA	
CUIR	Candidatura Unitaria De La Izquierda Regionalista												
CUP	Unidad Popular												
CUPS	Unitat Popular Per Socialisme												
DCAR	Democracia Cristiana Aragonesa											DCAR	
DCV	Democracia Cristiana Vasca												
DIV	Democratas Independientes Vascos												
DSCC	Democracia Social Cristiana De Cataluña												
EC.FED	Coalición Electoral Esquerra De Cataluña												

Fig. 1 Example of excerpt of metadata contained in a CANDIDADURAS spreadsheet. The first column shows the acronym of the candidacy. The second column offers the full name of the candidacy. The following columns give the acronyms used by each candidacy in each territory. In the case that a candidacy does not compete in a territory the corresponding cell is left blank.

has been included in each workbook to inform users that the data cannot be previewed in dataverse, indicating the need to download the files beforehand. The spreadsheets that can be found in each workbook, and an explanation of their content, are detailed below.

- **LEEME** (Readme, in Spanish): This spreadsheet indicates the primary source of the raw, provisional and official data, as well as the formula proposed to cite the data contained in the file. Given that the effort to homogenize, maintain and feed this electoral repository is onerous, in exchange for offering the data openly and free of charge, we at GIPEyOP ask all those analysts and researchers who make use of any of the databases to cite us in their publications.
- **LICENCIA** (Licence, in Spanish): This spreadsheet collects information on the license of use of the data. License CC-BY 4.0.
- **CANDIDATURAS** (candidacies, in Spanish): This spreadsheet provides details of all the candidacies that participated in the election referred to in the workbook and whose voting figures can be found in the results spreadsheets. This spreadsheet contains information presented in at least two columns, with the acronym of the candidacies in the first column and the name of the candidacy in the second. If the electoral space is divided into constituencies, the following columns, headed by the constituency identifiers (for example, INE code of province and name of province), contain the information with the acronym with which each candidacy competed in each constituency: see the example in Fig. 1.
- **MESAS** (Ballot boxes, in Spanish): This spreadsheet offers, among other variables, the results recorded at each ballot box.
- **SECCIONES** (Census sections, in Spanish): This spreadsheet offers, among other variables, the results recorded in each census section, obtained directly from the source when there is no data from ballot boxes or by aggregating the results by ballot boxes that make up each section.
- **DISTRITOS** (Local districts, in Spanish): This spreadsheet offers, among other variables, the results recorded in each district, obtained by aggregating the results by section that make up each district.
- **MUNICIPIOS** (Municipalities, in Spanish): This spreadsheet offers, among other variables, the results recorded in each municipality, obtained directly from the source when more disaggregated data are not available or by aggregating the results by district that make up each municipality. In the case of municipal elections, this tab is divided into MUNICIPIOS (PROVISIONAL) and MUNICIPIOS (OFICIAL), corresponding to the provisional data registered in each municipality, obtained by aggregating the results by districts, and to the official data obtained from the official source.
- **PROVINCIAS (PROVISIONALES)** (Provinces, in Spanish - provisional results): This spreadsheet—available for the workbooks of the datasets of general, regional and European elections—offers, among other variables, the provisional results registered in each province, obtained by aggregating the results by municipalities that make up each province (constituency or island). In the event that the natural division of the electoral space is by constituency or island, this is the name given to the spreadsheet.
- **PROVINCIAS (OFICIALES)** (Provinces, in Spanish - official results): This spreadsheet, available for the workbooks of the datasets of general, regional and European elections (in the latter case only for 2019 and 2014), offers, among other variables, the official results recorded in each province. In the event that the natural division of the electoral space is by constituency or island, the information refers to that area and that is the name given to the spreadsheet. The data contained in these spreadsheets has been obtained, in the case of the general and European elections, from the website of the Ministry of the Interior of Spain and, in the case of the regional elections, from the Central Electoral Board.

- CCAA (PROVISIONALES) (Regions, in Spanish - provisional results): This spreadsheet, available for the workbooks of the datasets of general and European elections, offers, among other variables, the provisional raw results recorded in each autonomous community obtained by aggregating the provisional results by provinces (constituencies or islands) that make up each autonomous community.
- CCAA (OFICIALES) (Regions, in Spanish - official results): This spreadsheet, available for the workbooks of the datasets of general and European elections, offers, among other variables, the official results recorded in each autonomous community, obtained by aggregating the official outcomes by provinces (constituencies or islands) that make up each autonomous community.
- INCIDENCIAS (Notes, in Spanish): This spreadsheet collects the incidents found, and the decisions taken to correct the errors found, in the data associated with the corresponding database.

One particular characteristic of the CANDIDADURAS spreadsheet in the general election workbooks is that the party name of the same national candidacy may differ between provinces (see Fig. 1). Elections to the Congress of Deputies can be seen as 52 different elections, although strongly interdependent and correlated³⁴, with different candidates in each constituency. For example, the Partido Socialista Obrero Español (PSOE) is called in Catalonia PSC.PSOE (acronym for Partit dels Socialistes de Catalunya - PSOE) while in the Basque Country it is called PSE.PSOE (acronym for Partido Socialista de Euskadi - PSOE) but both would be identified on the results spreadsheets with the generic acronym PSOE.

The internal composition of Excel workbooks described in the previous paragraphs shows the general situation. Later on in the paper, we specify graphically the level of information available for each specific election, since the information has not always been preserved for the lowest levels of aggregation.

Variables available in the spreadsheets of election outcomes. The number of variables (columns with information) available on each results spreadsheet is very heterogeneous, as it depends on the election, the geographical area to which it refers and, above all, the number of candidacies that competed. For some spreadsheets the number of available variables can exceed one hundred, while for others it may be less than twenty. Each row contains data of a voting unit and each column a variable, being votes to candidacies the more popular variables.

We now focus on briefly describing the variables that are present in almost all spreadsheets that contain poll results. Common variables that can be divided into two subsets: Identifier variables and Result (or numeric) variables.

Identifier variables. Each electoral process is held at a moment in time and in a specific place. The identifying variables are variables that collect this information, offering context and enabling the spatial and temporal location of the values associated with other variables. Identifier variables can be divided into temporal variables and spatial variables.

Temporal variables include ANYO, which identifies the year in which the election is held, and MES, which identifies the month of the electoral process. The values of these variables are common in all the rows of the same spreadsheet. The spatial variables are used to identify each polling unit, using the names (code) given to them by the National Institute of Statistics (INE).

The organization of elections is a logistically complex process that requires the intervention of a multitude of agents and must ensure that citizens with the right to vote know where to go to be able to exercise it. The INE is in charge of this task in Spain, informing each citizen with the right to vote of the ballot box (voting table) assigned to him/her and the place (polling station) where this ballot box is located. Each voter is assigned to only one polling station, where he/she can exercise the right to vote on election day.

The voting tables (ballot boxes) are decided as a consequence of dividing the voters registered in the same census section based on the first letter of their first surname. The census sections are created (directly or indirectly) by the INE after dividing (disjointedly) each municipality into small areas which vary in surface area but which contain, as a rule, a maximum of 1500 people with the right to vote in general elections. The census sections of the same municipality can, in turn, be grouped into districts.

The entire Spanish territory is hierarchically divided geographically into autonomous communities (including autonomous cities), provinces, municipalities, districts and census sections. Each of these units is identified with a code and, in the case of the more established administrative units (regions, provinces and municipalities), also with a name. The hierarchically-ordered concatenation of these codes makes it possible to assign a unique code to each ballot box. Each results spreadsheet has a first coding variable (for ballot box, section, district, municipality and provinces spreadsheets, respectively, being denoted *cod.mesa*, *cod.sssc*, *cod.dist*, *cod.mun* and *cod.prov*) that uniquely identifies each row of the corresponding spreadsheet. These unique codes are constructed as a concatenation of identifying codes for each of the higher scopes in the hierarchy (see Table 1).

For example, *cod.mesa* is an alphanumeric code of 13 characters resulting from the concatenation of the code (of the Ministry of the Interior) of the autonomous community (2 digits), and of the INE codes of province (2 digits), municipality (3 digits), district (2 digits), section (3 digits) and ballot box (1 letter). Each ballot box, in each election, therefore has a unique alphanumeric code that identifies it as well as the territorial units *included* in this code: census sections, districts, municipalities, provinces and autonomous communities. It should be noted, however, that in the case of databases corresponding to regional elections, the autonomous community code is omitted.

In addition to the unique identifier code resulting from the concatenation, the rest of the spatial identifier codes and names are also provided to help locate the data in the corresponding row of the data page. We are particularly interested in including the names of the geographical areas. Table 1 presents the description of the spatial variables.

Variable	Description
COD.CCAA	2-digit code identifying the Autonomous Community. The codes used are those of the Ministry of the Interior; INE codes differ from these for some Autonomous Communities.
CCAA	Denomination of the Autonomous Community
COD.PROV	2-digit INE code identifying the Province.
PROVINCIA	Denomination of the Province
COD.MUN	INE code of 3-digits identifying the Municipality within the province. The code 999 identifies a virtual municipality that includes absent residents with voting rights in the province.
MUNICIPIO	Denomination of the Municipality. Obviously this variable, and the previous one, are not available for spreadsheets that offer results at levels of aggregation higher than municipality level.
DISTRITO	2-digit identifier code of the District within the municipality. Obviously this variable is not available for spreadsheets that offer results at levels of aggregation higher than district level.
SECCIÓN	3-digit identifier code of the Census Section within the district. Obviously, this variable is not available for spreadsheets that offer results at levels of aggregation higher than the census section level.
MESA	Identifying letter of the ballot box within the census section. Obviously this variable is not available for spreadsheets that offer results at aggregation levels higher than the polling station level.

Table 1. Identifying variables of voting units. Source: compiled by the authors.

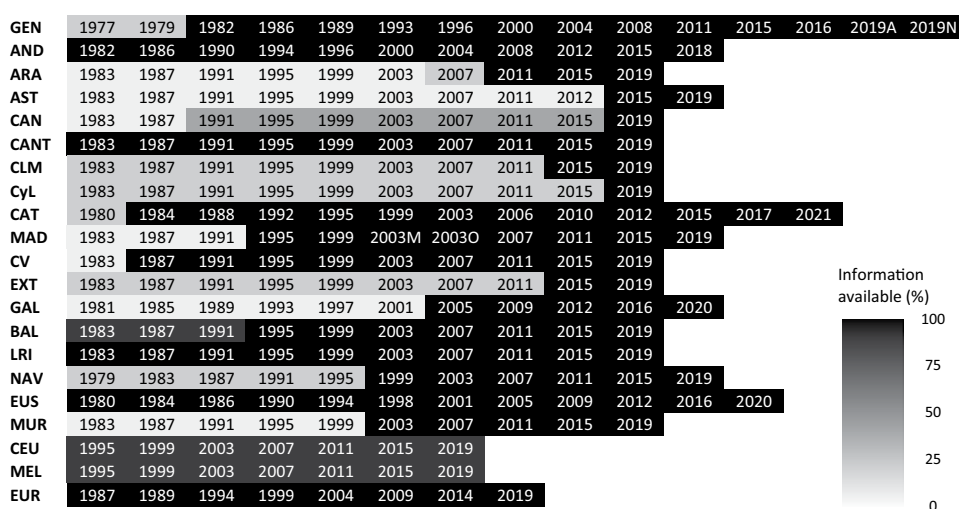


Fig. 2 Information available for national (General and European Parliament) and regional elections. The different levels of information available are represented using a descending scale of grey; the lower the tonality, the less information available, with black representing the maximum level of information. The acronyms to identify each of the electoral autonomous communities (cities) are provided in Online-only Table 1. The first row (GEN) refers to general elections, and the last row (EUR) to the European Parliamentary elections. The numbers represent the year that the elections were held.

The level of aggregation available for each of the general elections is detailed in the first row of Fig. 2. Specifically, for all general elections, except those of 1977 and 1979, data are available from ballot box level (the minimum level of disaggregation possible). For 1977 and 1979, data are available from the municipal level. The remaining rows in the figure report on the levels of detail available in each of the regional elections and the last row in the elections to the European Parliament.

Results or numerical variables. The temporal and spatial identifying variables tell us the time and the voting unit to which each of the values offered in the Results or Numerical variables corresponds. Table 2 provides a description of the main Results variables.

For the variables associated with votes for candidacies, the criteria of leaving blank cells has been used in the databases when a certain candidacy did not compete in a certain voting unit and, therefore, could not be voted on. This is different from a zero appearing. If a zero appears in the cell, it means that this candidacy did not obtain any vote in that spatial unit, even though it could have been voted on. The columns of candidacies are arranged in alphabetical order, or from highest to lowest number of votes in the election.

Other variables included in the databases. The previously mentioned variables are not the only ones contained in the databases. Depending on the election, in addition to these variables, there are additional variables (both identifying and numerical) that provide complementary and useful information for other analyses. For example, within the identifying variables, we find data spreadsheets that include variables such as CP (the postal code

Variable	Description
CENSO	INE Census. Number of resident electors with the right to vote.
CENSO + CERA	INE Census + CERA census (Absent Residents Electoral Census): Total number of voters with the right to vote.
CENSO.CERE	Census CERE. (Electoral Census of Foreign Residents).
VOT.AVANACE1	First advanced count of voter participation.
VOT.AVANACE2	Second advanced count of voter participation.
BLANCOS	Number of blank votes registered.
NULOS	Number of invalid (null) votes registered.
VOTOS.CANDIDATURAS	Sum of the votes received for the total number of candidacies.
CANDIDATURA_1	Number of votes received by candidacy 1. The corresponding initials appear as the name of the candidacy in each database.
CANDIDATURA_i	Number of votes received by the <i>i</i> -th candidacy. The corresponding initials appear in each database as the name of the candidacy.
CANDIDATURA_n	Number of votes received by the <i>n</i> -th candidacy. The corresponding initials appear in each database as the name of the candidacy.

Table 2. Results or Numerical variables. Source: compiled by the authors.

where the voting place, sometimes referred to as polling station, is located), COD.COLE (the code that the INE assigns to the voting place), COLEGIO (the name of the voting place), DIRECCION (the postal address, street and number, of the voting place) or variables such as LETRA_INI and LETRA_FIN, which identify the initial and final letters of the surnames of the people with the right to vote at each table voting, or ISLA or COMARCA, which identify the island or region where the voting unit to which the row refers is located.

Among the additional result variables, we can cite variables such as VOTOS.INTERVENTORES (votes of representatives of political parties not included in the census where they vote, but with the right to vote at the voting table where they are acting as representatives of the candidacies), as well as other variables such as: VOTOS.TOTALES (total of registered votes), VOTOS.ELECTORES (registered votes of voters included in the census lists), ABSTENCIONES (number of registered voters who have not exercised their right to vote) or VALIDOS (valid votes).

Obviously, although the variables in this subsection provide less added value, they may be of interest for some analyses. For example, with the help of the variables COD.COLE and DIRECCION, the percentages of votes registered by polling station could be represented on a map and thus the spatial patterns revealed or geo-statistical techniques³⁵ could be applied. Likewise, the variables VOTO.AVANACE1 and VOTO.AVANACE2 could be used to generate predictions of participation at the end of election day, which could be useful in an exit poll³⁶.

More details about the content of the datasets. In the previous sections we have given an overview of the content of the workbooks, mainly focused on the “Spanish General Elections” data set. Next, we describe, in different subsections, differentiating details regarding other data sets.

Regional elections. As a rule, the spreadsheets that define the Excel files for each regional election are similar to those for a general election, excluding spreadsheets CCAA (PROVISIONAL) and CCAA (OFFICIAL). There are some exceptions regarding the names of the spreadsheets or their content. Such is the case of Asturias where, instead of provinces, we speak of constituencies (Centre, West and East), and of the Balearic Islands, where the ISLAS spreadsheet refers to Formentera, Ibiza, Mallorca and Menorca, and of the Canary Islands, where the ISLAS spreadsheet contains information on El Hierro, Fuerteventura, Gran Canaria, La Gomera, Lanzarote, La Palma and Tenerife, grouped into two provinces: Las Palmas and Santa Cruz de Tenerife. The Canary databases, therefore, also include the PROVINCES spreadsheets.

Regarding the depth and level of information that can be found for each of the Autonomous Elections, we also observe a high degree of heterogeneity. As can be seen in Figs. 2 and 3, there are different degrees of information available. The most comprehensive corresponds to a situation where data is available at the ballot box level and, in addition, there is information on CERA voters. In the event that this latter information is available, the fictitious municipality that brings them together has been denoted in the databases by CERAx_x, where ‘xx’ represents the code of the corresponding province.

The information available can range from being complete to lacking certain data, such as, data on CERA voters or detailed results by ballot box, census section, municipality, or at worst, by (constituency) province level. The indicators on the levels of information associated with each election, detailed in Figs. 2 and 3, have been calculated as a percentage of the number of rows available in the databases and the potential total of rows. That is, the number of rows that the databases should contain if all ballot boxes reported their votes. Figure 2 provides detailed information for each election. Figure 3 presents, geographically, the global levels of information available from regional elections.

As can be seen, the general tonality of the map in Fig. 3 is not as dark as we would like, since only 7 of the 19 regions (36.8%) have (almost) complete information. The good news is that, at least for the last election, all the autonomous communities have information at ballot box level (see Fig. 2), a positive response that we hope will be maintained in successive elections.



Fig. 3 Information available in regional elections. The different levels of information are represented using a descending scale of grey; the lower the tonality, the less information available, with black representing the maximum level of information. The levels of information in each region have been calculated as a percentage of the number of rows available in the data spreadsheets over the total potential of rows that should exist if data relating to all levels of the election were reported. Rounding to one decimal place, the information levels are: Andalucía 99.9%, Islas Baleares 99.2%, Cantabria 99.9%, Ceuta 97.8%, Melilla 97.3%, La Rioja 99.7%, País Vasco 99.9%, Comunitat Valenciana 92.1%, Catalunya 93.6%, Comunidad de Madrid 74.9%, Navarra 61.5%, Murcia 53.2%, Galicia 45.6%, Canarias 34.4%, Aragón 31%, Castilla La Mancha 30.5%, Extremadura 29.6%, Castilla León 25.2% and Asturias 20.8%.

CCAA	Municipality
Andalucía	Jerez de la Frontera
Principado de Asturias	Gijón
Catalunya	Hospitalet de Llobregat Sabadell Terrassa
Galicia	Vigo
Comunidad de Madrid	Alcalá de Henares Fuenlabrada Getafe Móstoles
Región de Murcia	Cartagena
Comunitat Valenciana	Elche

Table 3. Downloadable information on local elections for large non-capital municipalities. Source: compiled by the authors.

Local elections. Spain currently has a total of 8,131 municipalities. In our opinion, attempting to offer detailed, user-friendly information files on each of the elections in each of the municipalities would be impractical. Obviously, this limitation could be remedied by developing a query application that allows the construction, on demand and in real time, from ASCII files of the files requested by an analyst. This application is still being developed, which is part of our future work plan, but in the meantime, in the “Spanish local elections” data set, we have decided to offer already prepared information on the results of local elections for the municipalities that would foreseeably arouse more interest. This includes a total of 52 provincial capitals and 2 CCAA capitals, Mérida and Santiago de Compostela, and 14 of the most populated large non-capital municipalities, excluding those already mentioned, as shown in Table 3.

The variables, as well as the spreadsheets included in each of the workbooks, are the same as in the other elections, with the exception that here the last spreadsheets refer to the municipality; provisional and official.

European elections. In the European Parliamentary elections, the results spreadsheets of the Excel workbooks cover the range from ballot box to the geographical area of the CCAA and are offered, as mentioned in Methods, in two versions. One file shows the votes of the candidacies presented and another with the parties that are part of the same national coalition. In the merged data workbooks, the CANDIDADURAS spreadsheet has a structure similar to that of a general election (see Fig. 1). As an example, Table 4 shows the candidacies for the 1987 elections that were presented under the same joint denomination with different names in different regions.

In the LEEME (README) spreadsheet of each of the merged files, the interested reader can consult the number of candidacies with different names that competed in the corresponding election and the number of groups of candidacies resulting from the merge of candidacies. Table 5 summarizes that information for all elections.

Group	Names of Candidacy
Coalición Izquierda de los Pueblos (CIP)	Esquerra dels Pobles-Entesa dels Nacionalistes (CIPENE), in Catalunya. Euskadiko Ezquerria-Izquierda de los Pueblos (EE.CIP), in Navarra and País Vasco. Esquerda dos Pobos PSG-EG (PSG.EG.CIP), in Galicia. PSM - Esquerra Nacionalista-Esquerria dels Pobles (PSM.EN.CIP), in Illes Balears. Coalición Izquierda de los Pueblos (CIP), in the rest of regions.
Izquierda Unida (IU)	Izquierda Unida-Iniciativa per Europa (IU.IPE), in Catalunya. Coalición Izquierda Unida (IU), in the rest of regions.
Por la Europa de los Pueblos (PNG)	Herrien European Alde-Europa de los Pueblos (EA.PNG.ERC) in País Vasco and Navarra. Per l'Europa de les Nacions (ERC.EA.PNG), in Comunitat Valenciana. Pola Europa dos Pobos (PNG.EA.ERC), in Galicia. Coalición por la Europa de los Pueblos (EA.ERC.PNG), in the rest of regions.
PSOE	Partit dels Socialistes de Catalunya (PSC.PSOE), in Catalunya. Partido Socialista Obrero Español (PSOE), in the rest of regions.

Table 4. Example. Groupings of candidacies in the European Elections of 1987. Source: compiled by the authors.

Year	Initial number of Candidacies	Number of groups
1987	44	35
1989	64	33
1994	58	35
1999	74	36
2004	72	31
2009	61	35
2014	70	39
2019	58	32

Table 5. Number of Candidacies and groups in European Parliamentary elections. Source: compiled by the authors.

Other files. Within the “Others” data set, sets of data are offered resulting from linking the electoral results with other types of information (for example, demographic or cartographic). Depending on the nature of the data, these are offered in Excel workbooks, which follow the same logic as indicated in the previous cases, or in other types of files. For example, in “Others” you can find the electoral cartography (for the scope of the census section) corresponding to the results of the elections to the Congress of Deputies of April 2019, and the results of the elections to the Basque Parliament (Eusko Legebiltzarra) held in July 2020. Such cartographies are offered in compressed files (rar) that contain all the geographic files necessary for their processing with appropriate software (e.g., ArcGIS, gvGIS, QGIS or R). The corresponding electoral results have been added to these files, through the associated dbf file. The LEEME.txt file contained in the rar files provides more details. Figure 4 shows the graphic representation of the participation rate registered in the two elections indicated above.

Due to its characteristics, the “Others” data set is the one with the greatest potential for future growth. In this sense, SEA is, and always will be, an unfinished project, as the possibilities of linking electoral results with other data sources are practically limitless. In terms of basic information, as well as adding to SEA the databases for the future electoral processes that will be held in Spain, one of our objectives is to complete those databases for which we do not have information for the most detailed levels of aggregation. Likewise, “Others” will also incorporate new workbooks obtained after matching results corresponding to different elections, estimating vote transfer matrices using ecological inference techniques, or associating new socioeconomic variables with electoral results. For the first task, the strategies proposed in Pavía and López-Quilez³⁷ and in Pavía and Cantarino³⁸ may be useful; for the second, the functions proposed in Lau *et al.*³⁹ and in Pavía and Romero⁴⁰; and, for the third, there are a multitude of socioeconomic variables, available at the census section level (such as income indicators, offered by the INE on its website within its section ‘Atlas of household income distribution’) which could be linked to election results. Another area of work on our agenda is related to obtaining electoral results prior to the Civil War. The provision of digitized data from that time would open up new study possibilities for interested historians.

Technical Validation

SEA contains the results of provisional counts (usually available from small areas) and official results (only available at the aggregate level) of all the electoral processes held in Spain after the restoration of democracy. As validation we have checked the concordance (within a minimum margin of error) between the aggregations of the provisional results and the official ones.

Sometimes the provisional counts do not match the official ones. This can occur naturally for two reasons. Firstly, because the provisional files do not have the CERA vote (census of absent Spanish residents). Secondly, because there have been challenges to the results initially recorded, during the election night, that are later resolved and never corrected in the provisional scrutiny files. Further anomalies may arise due to a correction that we implement from GIPEyOP when faced with inconsistencies between variables or due to temporal inconsistencies observed for certain outcomes. Such interventions are collected and documented in the INCIDENCIAS spreadsheet; see example in Table 6.

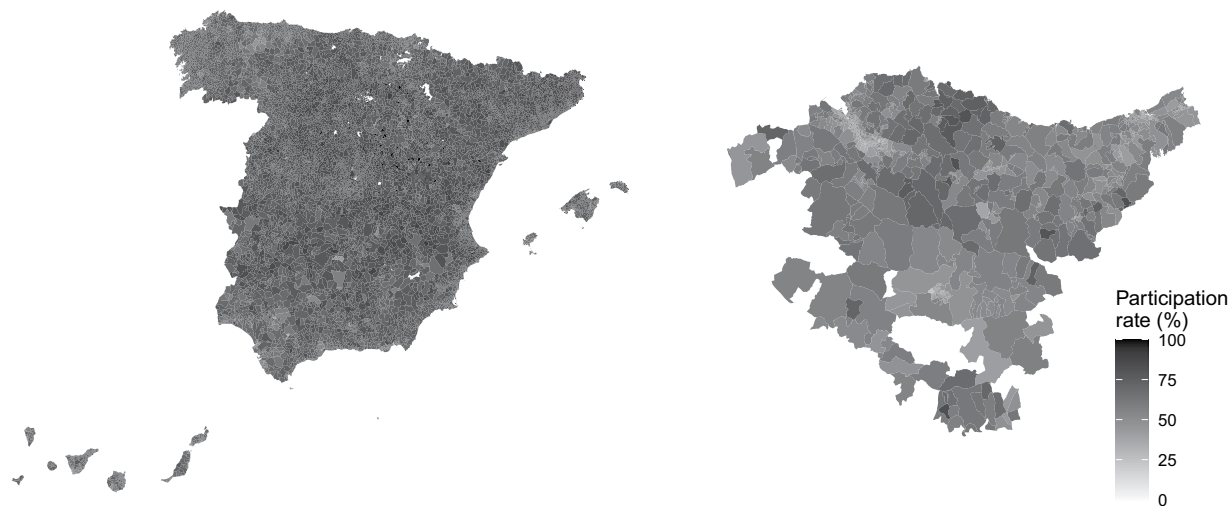


Fig. 4 Cartography of the participation rate in the elections to the Congress of Deputies of April 2019 (left panel) and in the elections to the Basque Parliament of July 2020 (right panel).

Error 1	In province 11 (Cádiz), municipality 001 (Alcalá de los Gazules), district 01, section 001, ballot box B, in the CENSO + CERA column the value 590 appeared, with 295 being the value associated with CENSO.
Action taken	After detecting that the census had doubled, the CENSO + CERA column was assigned the value of 295.
Error 2	In province 22 (Aragón), municipality 235 (Torres de Alcanadre), district 01, section 001, ballot box U, the CENSO + CERA column had the value 176, with 88 being the value associated with CENSO.
Action taken	After detecting that the census had doubled, the value 88 was assigned to the CENSO + CERA column.

Table 6. Example of data curation. General elections, April 2019 (in Spanish in the original). Source: compiled by the authors.

Given that the data offered in SEA correspond to the official provisional results, any modification of the raw data must be properly documented so that it can be reversed. We know that there are errors and misprints in the data, generally caused by small operational errors that any human activity entails. But these are often transferred to the results and proclaimed as final. The lack of incentive by political parties to request the revision of some misprints, even though they are aware of it, is often due to the fact that a correction may not have any practical consequence in terms of electoral results (distribution of representatives) and would have costs in terms of the legitimacy of the process.

The incidences that can be found in the databases are very varied. In our previous work, before considering undertaking the systematic organization of electoral results that SEA represents, we came across situations in which detected misprints persisted and were transferred to the official results. One example was seen in the results from a ballot box in which the votes received by the candidacies were shifted one column from a certain candidacy, as a consequence of a repetition or an omission counted as a zero. In these circumstances, the error can be dealt with in one of two ways; either leave the data as they are since they are official and have been proclaimed by the electoral authority as final, or correct the errors detected and document them so that the analyst can decide whether to work with the ‘correct’ data or with the official ones. At GIPEyOP we adopt this second approach.

Our future areas of work include systematically auditing all the existing databases in order to detect, correct and document all the errors and misprints that still exist in the provisional counting files. Thus, in the interest of this research group, any researcher who makes use of the files and detects any type of error should notify us so that we can reflect it as an incident. A very efficient way to detect errors in the data is by using the data. This will allow us all to have the best possible data to address future research. Documentation of the incidents detected and the actions implemented (see example in Table 6) is also essential. Having all the changes that are made to the ‘official’ data correctly documented is essential for allowing any changes introduced to be reversed, if deemed appropriate.

Usage Notes

When using the SEA database or part of it, please cite this manuscript and the particular references related to the workbook being used. For any questions, suggestions or requests for collaboration regarding SEA please contact the corresponding author.

Code availability

Processing of raw data has been performed using ad-hoc scripts in the statistical software R, version 4.0.2 (R Core Team, 2020). Code is available on http://gipeyop.uv.es/gipeyop/base_datos/21_otros/CODIGO_SEA.zip.

Received: 2 March 2021; Accepted: 18 June 2021;
Published online: 28 July 2021

References

- Boletín Oficial del Estado. *Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno*. BOE, **295**, 10 de diciembre (2013).
- Mora-Navarro, G., Balaguer-Beser, A., Martí Montolí, C. & Femenia-Ribera, C. Geoportal for electoral geomarketing to detect microzones with potential voters in an urban area. *Mapping* **29**, 34–43 (2020).
- Nello, O. & Gomà, A. Geographies of discontent: urban segregation, political attitudes and electoral behaviour in Catalonia. *City, Territory and Architecture* **5**, 23 (2018).
- Pavía, J. M. & Larraz, B. Nonresponse bias and superpopulation models in electoral polls. *Revista Española de Investigaciones Sociológicas* **137**, 237–264 (2012).
- Scheuren, F.J. & Alvey, W. *Elections and Exit Polling*. (Hoboken: Wiley, 2008).
- Pavía, J. M. Improving predictive accuracy of exit polls. *International Journal of Forecasting* **26**, 68–81 (2010).
- Pavía, J. M., Badal, E. & García-Cárceles, B. Spanish exit polls: sampling error or nonresponse bias? *Revista Internacional de Sociología* **74**, e043 (2016).
- Pavía, J. M. & Larraz, B. Quick-counts from non-selected polling stations. *Journal of Applied Statistics* **35**, 383–405 (2008).
- Pavía, J. M., Larraz, B. & Montero, J. M. Election forecasts using spatiotemporal models. *Journal of the American Statistical Association* **103**, 1050–1059 (2008).
- Klima, A., Schlesinger, T., Thurner, P. W. & Küchenhoff, H. Combining aggregate data and exit polls for the estimation of voter transitions. *Sociological Methods & Research* **48**, 296–325 (2019).
- Pavía, J. M. & Aybar, C. La movilidad electoral en las elecciones de 2019 en la Comunitat Valenciana. *Debats. Revista de Cultura, Poder y Sociedad* **134**, 27–51 (2020).
- Romero, R., Pavía, J. M., Martín, J. & Romero, G. Assessing uncertainty of voter transitions estimated from aggregated data. Application to the 2017 French presidential election. *Journal of Applied Statistics* **47**, 2711–2736 (2020).
- Bernardo, J. M. Monitoring the 1982 Spanish socialist victory: A bayesian analysis. *Journal of the American Statistical Association* **79**, 510–515 (1984).
- Pavía-Miralles, J. M. Forecasts from non-random samples: the election night case. *Journal of the American Statistical Association* **100**, 1113–1122 (2005).
- O'Loughlin, J., Flint, C. & Anselin, L. The geography of the nazi vote: context, confession, and class in the Reichstag election of 1930. *Annals of the Association of American Geographers* **84**, 351–380 (1994).
- Mora, J. La abstención electoral en España. Un enfoque espacio-temporal mediante datos agregados. Master Thesis. Universidad de Valencia (2020).
- Altman, M. & McDonald, M. P. BARD: Better automated redistricting. *Journal of Statistical Software* **42**, 1–28 (2011).
- Fifield, B., Higgins, M., Kosuke, I. & Tarr, A. Automated redistricting simulation using Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics* **29**, 715–728 (2020).
- R Core Team. *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/> (2020).
- RStudio Team. *RStudio: Integrated Development for R*. <http://www.rstudio.com/> (2020).
- Boletín Oficial del Estado. Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General. BOE, **147**, 20 de junio (1985).
- Pavía, J.M. & Toboso, F. Key electoral institutions and rules influencing proportionality and partisan bias in Spanish politics. *Panaeconomicus*, Online available, 1–22, <https://doi.org/10.2298/PAN180725022P> (2021)
- Pavía, J. M. & Veres, E. Un nuevo estimador para disgregar totales poblacionales. El caso de los nuevos electores. *Anales de Economía Aplicada* **30**, 817–826 (2016).
- Pavía, J.M. & Veres, E. Desagregando estadísticas de población (In Herrerías, J. M.; Callejón, J.) *Investigaciones en Métodos Cuantitativos para la Economía y la Empresa*. Granada: Universidad de Granada, 543–555 (2016).
- Pavía, J. M. & Cantarino, I. Can dasymetric mapping significantly improve population data reallocation in a dense urban area? *Geographical Analysis* **49**, 155–174 (2017).
- King, G. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. (Princeton: Princeton University Press, 1997).
- Pavía, J. M. & García-Cárceles, B. Una aproximación empírica al error de diseño muestral en las encuestas electorales del CIS. *Metodología de Encuestas* **14**, 45–62 (2012).
- Evans, J. & Ivaldi, G. Contextual effects of immigrant presence on populist radical right support: testing the 'halo effect' on FN voting in France. *Comparative Political Studies*, online available, 1–32, <https://doi.org/10.1177/0010414020957677> (2020).
- Pavía, J. M., Aybar, C. & Pérez, V. Spanish General Elections. *SEA Database. Harvard Dataverse* <https://doi.org/10.7910/DVN/0GPRYW> (2021).
- Pavía, J. M., Aybar, C. & Pérez, V. Spanish Regional Elections. *SEA Database. Harvard Dataverse* <https://doi.org/10.7910/DVN/BJRXXX> (2021).
- Pavía, J. M., Aybar, C. & Pérez, V. Spanish Local Elections. *SEA Database. Harvard Dataverse* <https://doi.org/10.7910/DVN/53FCE6> (2021).
- Pavía, J. M., Aybar, C. & Pérez, V. Spanish European Elections. *SEA Database. Harvard Dataverse* <https://doi.org/10.7910/DVN/BXSFUF> (2021).
- Pavía, J. M., Aybar, C. & Pérez, V. Spanish Elections - Others. *SEA Database. Harvard Dataverse* <https://doi.org/10.7910/DVN/5MMHUI> (2021).
- Pavía, J. M., García-Cárceles, B. & Badal, E. Estimating representatives from election poll proportions: The Spanish case. *Statistica Applicata – Italian Journal of Applied Statistics* **25**, 325–340 (2016).
- Pavía, J. M., Montero, J. M. & Larraz, B. Geografía, elecciones y predicciones electorales. *Investigación y Marketing* **98**, 70–77 (2008).
- Sobrino, V. M. Los avances de resultados en una jornada electoral. *Más Poder Local* **13**, 42–44 (2012).
- Pavía, J. M. & López-Quilez, A. Spatial vote redistribution in redrawn polling units. *Journal of the Royal Statistical Society, Series A – Statistics in Society* **176**, 655–678 (2013).
- Pavía, J. M. & Cantarino, I. Dasymetric distribution of votes in a dense city. *Applied Geography* **86**, 22–31 (2017).
- Lau, O., Moore, R.T. & Kellermann, M. *eiPack: Ecological Inference and Higher-Dimension Data Management*. R package version 0.2-1, <https://CRAN.R-project.org/package=eiPack> (2020).
- Pavía, J. M. & Romero, R. Improving estimates accuracy of voter transitions. Two new algorithms for ecological inference based on linear programming. *Advance. Preprint*. <https://doi.org/10.31124/advance.14716638.v1> (2021).

Acknowledgements

The authors wish to thank Marie Hodkinson for translating the text of the paper into English, the Editorial Board Member Mercè Crosas for her first-rate assistance answering all our queries, including issues about the submission process and where to deposit our data sets and two anonymous reviewers for their valuable suggestions. The authors acknowledge the support of Generalitat Valenciana through project AICO/2021/257 (Conselleria d'Innovació, Universitats, Ciència i Societat Digital).

Author contributions

This publication is the result of the joint work of the three authors over several years. Although all three authors have been involved, to a greater or lesser extent, in the different phases of the project, V.P. focused on the creation and curation of the data and contributed to the writing of the paper; C.A. provided raw data, created and curated the data and contributed to the writing of the paper; J.M.P. designed the research, provided raw data and contributed to the writing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.M.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2021

Anexo II

Expanding the New Roles of the Military – The Case of Spain’s Military Emergency Unit: A Research Note

Pérez, V., Aybar, C. & Pavía, J. M. (2022). *Armed Forces & Society*, 0(0).

<https://doi.org/10.1177/0095327X221130360>

Expanding the New Roles of the Military—The Case of Spain’s Military Emergency Unit: A Research Note

Armed Forces & Society

1–12

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0095327X221130360

journals.sagepub.com/home/afs



Virgilio Pérez¹, Cristina Aybar¹, and Jose M. Pavía¹

Abstract

This note illustrates some of the new roles that armed forces are taking on in developed countries, highlighting the structure, capabilities, and missions developed by the Spanish Military Emergency Unit (UME). The note describes some of these new roles assumed by UME, such as its response to the COVID-19 pandemic, and shows, using official temporal and georeferenced data, the growing use of the military to support civilian authorities in response to natural disasters, man-made disasters, and other emergencies alike. In addition, we also analyze public opinion survey data to argue that this is helping to improve society’s perception of the armed forces even in a country like Spain where, due to its recent history, a significant part of its population is reluctant to the military. In short, the military is taking on more emergency roles as needs arise and these new roles can help to improve its perception by citizens.

Keywords

military emergencies unit, civil protection, civil-military emergency preparedness, military assessment, Spain

The armed forces are an essential part of the organization of any state, their main task being to defend the country against external threats, current, or potential. Sometimes,

¹Research Unit in Economy of Defence, Cybersecurity and Protection, GIPEyOP, University of Valencia, Spain

Corresponding Author:

Jose M. Pavía, Research Unit in Economy of Defence, Cybersecurity and Protection, GIPEyOP, University of Valencia, 46022 Valencia, Spain.

Email: pavia@uv.es

however, the armed forces also perform functions that go beyond their constitutional mandates as a coercive tool to promote and protect the country's interests abroad (Edmunds, 2006). To these tasks, in recent years, others have been added related to the (re)construction of infrastructures, security or training in post-war scenarios abroad, as well as a growing involvement in homeland civil protection actions. In many countries, certain military actions in non-war contexts have caused a unique, complex and ambiguous (sometimes even contradictory) relationship between civilian and military elites (Burk, 2002). In general, the military authorities try to reconcile the traditional tasks, which establish the preparation for a potential war scenario, with the sense of duty, inherent to the call of society.

Natural disasters (major forest fires, floods, earthquakes, hurricanes, volcanic eruptions, . . .), increasingly devastating phenomena, allow the armed forces to collaborate with civil authorities, contributing their resources and capabilities with the aim of mitigating the effects on the population of an emergency or catastrophe (Heaslip & Barber, 2014). This contribution is made, fundamentally, by complementing the national civil protection services managed by regional and local authorities, and participating only in those cases in which the magnitude of the catastrophe exceeds civil capacities (Cabigiosu, 2005). Although the armed forces usually enjoy considerable legitimacy in society (Ramalho, 2018), interventions of this nature, as well as other subsidiary actions, such as the so-called "civic-social" activities, represent opportunities to reinforce the image of the armed forces in society (Deverell et al., 2015).

The characteristics inherent in the capacity, organization and comportment of the military have resulted in various countries in recent years opting to specialize, within their armed forces, certain units to deal effectively with these national emergency situations (Goniewicz et al., 2019). The United States, Canada, France, Japan, or Switzerland are just a few examples of nations that perceived the need to assign their military-specific functions related to emergencies and civil protection (Roldán Pascual, 2012), not only because of the urgent and unforeseen circumstances that could arise, but also because of the proven risk involved in the intervention of non-expert personnel in these circumstances. Prominent examples of military actions in emergencies include Hurricane Katrina (a natural disaster that mobilized 70,000 military personnel) (Burke, 2016) and the 2011 Great East Japan Earthquake (and tsunami), which damaged the reactors of the Fukushima nuclear power plant and impelled the Japanese Self-Defence Forces to mobilize more than 100,000 soldiers, 540 aircraft, and 59 vessels (Takeshita et al., 2020). As an example, the main tasks of the French *Unités d'Instruction et Intervention de la Sécurité Civile* include to (1) fight forest fires; (2) fight against environmental pollution; (3) act against floods; (4) protect from technological risks (including CBRN); (5) search and rescue, particularly after earthquakes; and (6) logistical support in operations (Gallegos & Díez Alcalde, 2013).

The aim of this paper is to expound the additional new roles of civil protection that the armed forces are assuming in developed countries by showing the structure, capacities and missions of the UME. We conclude by arguing that these new roles can help to improve the perception of the military by the population.

The Military Emergency Unit

Despite the fact that more than 40 years have passed since the establishment of democracy in Spain, a part of Spanish society continues to see the armed forces as related to concepts such as war and conflict, and not to defense or protection. The Civil War and Franco's military dictatorship left a deep imprint of mistrust toward the army (Tur-Prats & Valencia Caicedo, 2020). Into this scenario was born the Spanish Military Emergency Unit (UME) during the first decade of the 21st century.

The Founding of UME

In the summer of 2005, a devastating forest fire caused by the carelessness of some hikers burned down more than 10,000 hectares of forests and crops in the province of Guadalajara (located in the center of the Iberian Peninsula). The fire also claimed the lives of 11 firefighters. A few months earlier, a severe ice and snow storm caused thousands of people to be trapped on the roads linking Madrid with the north of Spain. The magnitude of these events revealed the need for a unit that, at a national level: (1) had the capacity to face risks, calamities, or public misfortunes; (2) acted in emergency circumstances and provided collaboration and support to the regional authorities; (3) offered an energetic and effective response; and (4) guaranteed the safety of people when a catastrophe put them in serious danger.

These qualities perfectly reflect the military, which bring together a series of attributes—personal and material—that make them uniquely suitable for reacting quickly and effectively in stressful situations. According to the Royal Ordinances (Boletín Oficial del Estado [BOE], 2009) for the Spanish Armed Forces (SAF), the military builds its comportment on discipline, hierarchy and unity, essential characteristics for achieving maximum efficiency from its actions, ensuring high levels of coordination. These are characteristics, which are absolutely essential in emergency situations. This, together with the ability to deploy quickly and in an orderly manner on the ground, moving both themselves as well as cumbersome equipment in a short time, make them a valuable service to society, cushioning the effects of catastrophes and instilling confidence in the civilian population. This argument, after extensive deliberation not without controversy, led to the creation on October 7, 2005, by agreement of the Council of Ministers, of the UME (BOE, 2006).

The decision to create a specific and specialized military unit as a civil protection instrument sparked misgiving, with criticism from various spheres at the time of its creation, including from the military institution itself. Hence, only the seriousness of the situation, or its potential evolution, justifies military action in Spain and, if the circumstances arise, this request must be addressed to the Ministry of Defence. It must be remembered that the actions of the military continue to be questioned in Spanish society and, therefore, issues related to the military require careful consideration. Consequently, military support complements others that are already on the ground, with the civil authorities being responsible for the operational direction of the emergency.

UME Structure, Capabilities, and Mission

UME is a permanent combined force with the capacity to intervene anywhere in Spain or abroad. Its mission is to contribute to the safety and well-being of citizens in cases of serious risk, catastrophe, calamity or other public needs (BOE, 2019). The Unit has a total of 3,600 soldiers from the Army, Navy, Air Force, and Common Corps (Ministerio de Defensa, 2022). It has a Headquarters Unit (UCG) from where the five Emergency Intervention Battalions (BIEMs) are directed and controlled. Each BIEM has a particular area of action, a circumstance that allows for better knowledge of the terrain and closer contact with both the civil authorities and the personnel who work in the different institutions that act in emergencies. In addition to UCG and BIEMs, UME has a Transmissions Battalion (BTUME) and a Support Regiment (RAIEM), with access, when necessary, to personnel and resources from other Units of the armed forces to help in disaster management.

The most common emergencies that have required UME's response are forest fires, floods, heavy snowfall, and other adverse weather events. In general, UME goals are to mitigate the risk of man-made or natural disasters and to deal with threats and emergencies. Hence, there are other natural risks such as earthquakes, tsunamis, landslides, or volcanic activity that, although less frequent in Spain, are also on UME's radar. To the above list should be added the emergencies caused by technological and environmental risks, as well as by terrorist attacks or illegal and violent acts, and those against critical infrastructures or with chemical, biological, radiological, and nuclear (CBRN) agents. Similarly, cultural heritage, which represents the historical legacy and hallmark of the people that constitute Spain, is not free from the risk of damage as a result of catastrophes, whether from natural or human-induced disasters. In summary, the capabilities that UME can provide are: (1) command and control; (2) operational direction of emergencies of national interest; (3) intervention in forest fires; (4) intervention in floods; (5) intervention in winter storms; (6) intervention in earthquakes, volcanoes, and landslides; (7) intervention in technological risks (including CBRN); (8) environmental intervention; (9) attention to the affected and injured population; (10) protection of assets of cultural interest; (11) intervention in emergencies after terrorist attacks, or illicit and violent acts; (12) support to the State Security Forces and Bodies and other Public Administrations; (13) intervention in major air or rail accidents; and (14) attention to multiple victims.

UME Response to Emergencies

Since its creation in 2005 and up to 2021, UME has intervened on more than 600 occasions throughout the Spanish territory (see Figure 1, left panel), being a much-used support resource by all regional authorities, even by those that traditionally remain more distant from the military institution. Particularly, after collecting and analyzing official data from UME and National Forest Fire Information Coordination Centre (CCINIF, 2022; Ministerio de Defensa, 2022), one can see that actions in large forest fires, winter storms or floods have highlighted the usefulness and operability of this nascent military unit. But more than this, it has made the work of SAF more

visible, supporting and protecting Spanish citizens and strengthening ties between the military and society. Figure 1 (left) shows, by region, the actions performed by UME between 2005 and 2021 (Ministerio de Defensa, 2022), while Figure 1(right) presents its relative participation in large fires during 2007–2021 (CCINIF, 2022).

Although it is true that most of the emergency actions carried out by UME have been based on fighting forest fires (see Figure 2, left panel), the missions carried out in its 15-year history have been very diverse (Ministerio de Defensa, 2022), highlighting its versatility and adaptability. This circumstance is evident in the last 5 years in which UME has participated in actions of great social impact, such as the “Extremadura” Environmental Operation, in which it removed 170,000 tons of *Eichhornia crassipes*, an aquatic plant that had invaded more than 175 km of the riverbed of the Guadiana river and was causing disastrous economic and environmental damage.

In fact, the involvement of UME in catastrophes has been growing over time. In recent years, an increasing trend can be seen in the number of days in which the unit has acted (see Figure 2, right panel) (Ministerio de Defensa, 2022). Indeed, at the time of revising this paper (mid-August, 2022), UME had already performed more actions than ever in a single year (personal communication). This circumstance highlights the use that the civil authorities are making of UME, the latter showing itself to be a valuable element of support in critical situations for the civilian population.

The UME Actions During the COVID-19 Crisis

The public health emergency situation brought about by the COVID-19 pandemic caused the Government of Spain to take the decision on March 14, 2020, to declare a state of alarm throughout the country to deal with an unprecedented emergency situation. A day later, the first coordination meeting to address the measures to be implemented by SAF took place at the headquarters of the Ministry of Defence in which it was decided to activate UME to carry out reconnaissance in the main Spanish cities. These actions were the beginning of Operation Balmis, named after the military doctor and surgeon who brought the smallpox vaccine to the territories of the Spanish empire in America and the Philippines at the beginning of the 19th century.

The implementation of measures was carried out diligently and rapidly, a necessary response to the rapid spread of the SARS-CoV-2 coronavirus. At that time, neither of them knew what was really happening, nor could they quantify the risk that was being run. In this situation of fear and uncertainty (García-Real et al., 2020), citizens saw that, together with the health services and civil protection forces, the military were among the first to be at the frontline in tackling the pandemic. The operation helped to strengthen the image of the military (Hidalgo, 2021).

During the 98 days that Operation Balmis lasted, representing the greatest effort and military deployment carried out in peacetime in Spain, SAF completed 20,000 actions (5,300 disinfections in nursing homes, 4,800 actions to support hospitals and health centers, 1,200 patient transfers, 20 field hospitals built, etc.) in more than 2,300 municipalities, supporting health services and security forces in their efforts to contain the spread of

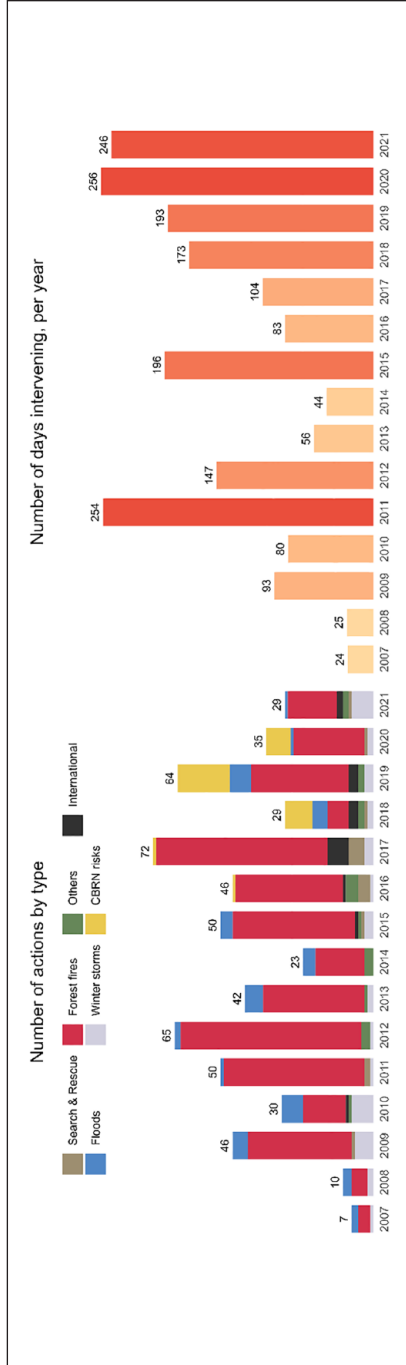


Figure 2. Number of Actions Carried Out by Military Emergency Unit (2007–2021), by Type (Left Panel) and Number of Days Intervening, Per Year (Right Panel).
 CBRN = chemical, biological, radiological, and nuclear risks.

the coronavirus (Fumero, 2021; Infodefensa, 2020). In addition to these tasks, they performed the difficult mission of helping to transfer the deceased to authorized facilities.

Operation Balmis was not the only military action to take place in Spain related to the COVID-19 pandemic. After a period of strict home confinement, the Spanish returned to the streets, maintaining certain hygienic-sanitary measures, but these measures were not enough to stop the spread of the virus and at the end of September 2020, before a new outbreak (Pérez et al., 2022), the military was activated again. This time, the military collaborated with the regional authorities in track and trace operations to locate and inform individuals of possible exposure to infection, a fundamental task in this phase of the pandemic. This operation named Mission Baluarte, and still ongoing at the time of the writing of this paper, treated the virus like an invisible enemy in an armed conflict (Expósito et al., 2020).

As in Operation Balmis, the first to track the disease were UME soldiers, which gave time for the other SAF units to acquire training and organize their supplies (Tarilonte, 2020), and for the health institutions of the different regions to establish action protocols. With this action, UME reaffirmed itself as an element of cohesion and solidarity among Spaniards.

Public Opinion and Armed Forces

This section presents some results obtained after analyzing public opinion data that could serve as a way of supporting the hypothesis for Spain that the actions of UME are contributing to improving the perception that Spaniards have of the military.

The Centro de Investigaciones Sociológicas (CIS), the most prestigious Spanish institution in the field of public opinion studies (Pavía & Aybar, 2018), offers the scientific community and society a database made up of almost 2,000 studies, carried out over the past 60 years, which enable an analysis of the behavior and opinion of society on multiple issues. An example of these studies is the one carried out between January 1997 and September 2017 under the title “National Defence and the Armed Forces” (CIS, 2022). A total of 12 surveys (identified with numbers 2234, 2277, 2317, 2379, 2447, 2592, 2680, 2825, 2912, 2998, 3110, and 3188 in the CIS database), with generally a biannual periodicity and $n \approx 2500$ as average sample size, enable an analysis of how the perception held by Spanish society of SAF has evolved in relation to several issues. Figure 3 shows the result regarding the military as a profession. As can be seen, the evolution of this variable shows a change in trend after the creation of the UME, which can be taken as an indication of the positive influence that this unit could be having on the public’s assessment of the armed forces. Causality, however, cannot be claimed.

The 12 surveys carried out by CIS enable us to analyze the period 1997–2017. In 2019, the Sociological Observatory on citizen perception of the activities, plans and programs of the Defence culture and awareness plan, backed by the Ministry of Defence, prepared a report based on a survey very similar to those carried out by CIS. The results of this report continue to reflect the growing trend of a positive assessment by Spaniards of their military, with the attached report recommending, based on the

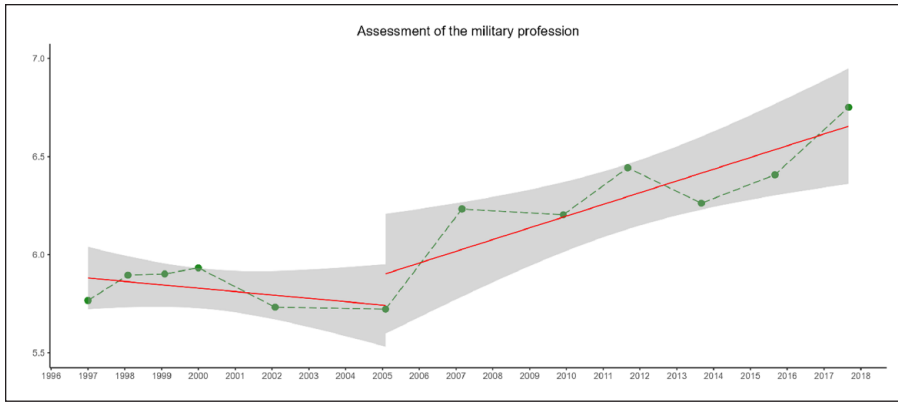


Figure 3. Assessment (Average) in the Period 1997–2017 of the Military Profession
 Note. The higher the number, the better the assessment.

opinion of those surveyed, a greater presence of SAF in society, not only in the areas of environmental catastrophes but also carrying out social and welfare work. In short, the presence of more UME-type units is making an impact (Ministerio de Defensa, 2019).

The latest available study that addresses issues on SAF and society in Spain was carried out in October 2021 (Cancio, 2021). This report shows the sympathy and recognition that Spanish society shows toward its armed forces, especially after the Balmis and Baluarte operations. In fact, 80% of those surveyed have a good or very good opinion of the military, a figure substantially higher than previous studies. However, previous comments aside, in no study is the respondent directly asked about the assessment of UME.

Discussion and Conclusions

Thousands of studies conducted by researchers around the world have recorded an increase in the frequency of natural disasters: a frequency that is expected to grow over the next few years, having a huge impact on the global economy and causing human and material damage in all parts of the world. This circumstance, along with the need to justify large budget allocations to defense departments, is pushing many developed countries to assign new social and civil protection roles to the armed forces.

The measures that are being implemented in different developed countries focus on reducing the damage that future catastrophes may cause to the civilian population, with each territory developing strategies adapted to its own circumstances. Fighting forest fires, emergency response to floods, winter storms or earthquakes, or containment/disinfection tasks against CBRN agents are just some of the many services that the armed forces make available to civil authorities. The assignments allocated to the armed forces are growing, broadening the concept of national defense to deal with situations that are quite distinct from armed conflicts.

In Spain, UME has been able to unify criteria and action protocols, demonstrating the value of one of the many qualities of the military: discipline. Operations such as Balmis and Baluarte, or others mentioned in this note, have led to UME (and by extension the armed forces) becoming an element of cohesion that transmits security and comfort to its citizens and generates confidence within the society in which it serves. The missions carried out by UME are growing diverse, COVID-19 actions are examples, with public opinion data pointing to the hypothesis that this also serves to improve population attitudes toward the armed forces, even in a country like Spain with a military dictatorship in its recent history. Future research should be tailored to investigate the trends in the Spanish citizens' perceptions of UME and the tasks that the latter performs. An ad hoc survey would make it possible to directly compare the perceptions that citizens have of UME and SAF.

Acknowledgments

The authors thank the chief editor and two anonymous reviewers for their valuable comments and suggestions. Thanks are also due to Marie Hodkinson for translating the text of the paper into English and the staff of the National Forest Fire Information Coordination Centre (CCINIF) for providing the microdata on forest fires in Spain (from 1996 to 2015). The usual disclaimers apply.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research has received the support of the Generalitat Valenciana through project AICO/2021/257 (Conselleria d'Innovació, Universitats, Ciència i Societat Digital) and of Ministerio de Ciencia e Innovación through project PID2021-128228NB-I00.

References

- Boletín Oficial del Estado. (2006). Resolución de 19 de enero de 2006, de la Subsecretaría, por la que se da publicidad al Acuerdo de Consejo de Ministros por el que se crea la Unidad Militar de Emergencias (UME) [Resolution of 19 January 2006, of the Undersecretary's Office, publicising the Agreement of the Council of Ministers creating the Military Emergency Unit (UME)]. *Boletín Oficial del Estado*, 17, 2593.
- Boletín Oficial del Estado. (2009). Real Decreto 96/2009, de 6 de febrero, por el que se aprueban las Reales Ordenanzas para las Fuerzas Armadas [Royal Decree 96/2009 of 6 February 2009, approving the Royal Ordinances for the Armed Forces]. *Boletín Oficial del Estado*, 33, 13008–13028.
- Boletín Oficial del Estado. (2019). Orden DEF/160/2019, de 21 de febrero, por la que se regula la organización y funcionamiento de la Unidad Militar de Emergencias [Order DEF/160/2019, of 21 February, regulating the organisation and operation of the Military Emergency Unit]. *Boletín Oficial del Estado*, 46, 17040–17047.

- Burk, J. (2002). Theories of democratic civil-military relations. *Armed Forces & Society*, 29(1), 7–29. <https://doi.org/10.1177/0095327X0202900102>
- Burke, R. (2016). Lessons from Katrina: Commanding the military during disaster response—then and now. *International Journal of Emergency Management*, 12(3), 221–240. <https://doi.org/10.1504/IJEM.2016.079016>
- Cabigiosu, C. (2005). The role of Italy’s military in supporting the civil authorities. *Connections*, 4(3), 59–82
- Cancio, F. (2021, October 12). Los españoles se sienten seguros con las Fuerzas Armadas [Spaniards feel safe with the Armed Forces]. *La Razón*. <https://www.larazon.es/espana/20211012/kwnj5fjzjenlp5w6xwmnekqxa.html>
- Centro de Investigaciones Sociológicas. (2022). *las Encuestas del CIS [The CIS Surveys]*. https://www.cis.es/cis/openncm/ES/1_encuestas/index.jsp
- Deverell, E., Olsson, E.-K., Wagnsson, C., Hellman, M., & Johnsson, M. (2015). Understanding public agency communication: The case of the Swedish armed forces. *Journal of Public Affairs*, 15(4), 387–396. <https://doi.org/10.1002/pa.1552>
- Edmunds, T. (2006). What are armed forces for? The changing nature of military roles in Europe. *International Affairs*, 82(6), 1059–1075. <https://doi.org/10.1111/j.1468-2346.2006.00588.x>
- Expósito, J. L., Tarilonte, E., & Hernández, V. (2020). Todos los días son lunes [Every day is Monday]. *Revista Española de Defensa*, 371, 10–24.
- Fumero, P. (2021, March 14). Zaforteza: “Balmis fue el mayor esfuerzo de las Fuerzas Armadas para salvar vidas” [Zaforteza: “Balmis was the greatest effort of the Armed Forces to save lives”]. *El Día: La Opinión de Tenerife*. <https://www.eldia.es/sociedad/2021/03/14/balmis-mayor-esfuerzo-fuerzas-armadas-42504976.html>
- Gallegos, J., & Díez Alcalde, J. (2013). Las Fuerzas Armadas de diferentes países en la Protección Civil [The Armed Forces of different countries in Civil Protection]. *Cuadernos de Estrategia*, 165, 181–214.
- García-Real, J., García-Silgo, M., Conejo-García, M., Samper-Lucena, E., Torrejón-Correa, C., Cerezo-Ureta, J., Más-Esquerdo, J., Laplaza-Brun, A., & Ruiz-Espinosa, J. (2020). La psicología militar en la prevención de salud mental durante la pandemia por SARS-Cov-2 [Military psychology in mental health prevention during the SARS-Cov-2 pandemic]. *Sanidad Militar*, 76(2), 118–125.
- Goniewicz, K., Goniewicz, M., & Burkle, F. (2019). The territorial defence force in disaster response in Poland: Civil-military collaboration during a state of emergency. *Sustainability*, 11(2), 487. <https://doi.org/10.3390/su11020487>
- Heaslip, G., & Barber, E. (2014). Using the military in disaster relief: Systemising challenges and opportunities. *Journal of Humanitarian Logistics and Supply Chain Management*, 4(1), 60–81. <https://doi.org/10.1108/JHLSCM-03-2013-0013>
- Hidalgo, M. (2021). El papel de las Fuerzas Armadas en la gestión de la COVID-19 como generador de confianza [The role of the Armed Forces in the management of COVID-19 as a generator of trust]. *Boletín Instituto Español de Estudios Estratégicos*, 21, 188–200.
- Infodefensa. (2020, June). Las cifras de la operación Balmis: 20.000 intervenciones en 98 días de lucha [The figures of the Balmis operation: 20,000 interventions in 98 days of fighting]. <https://www.infodefensa.com/texto-diario/mostrar/3126495/cifras-operacion-balmis-20000-intervenciones-98-dias-lucha>
- Ministerio de Defensa. (2019). *Informe Defensa 2019 [Defence Report 2019]*. <https://publicaciones.defensa.gob.es/>

- Ministerio de Defensa. (2022). *Unidad Militar de Emergencias* [Military Emergency Unit]. (website). <https://www.defensa.gob.es/ume/>
- National Forest Fire Information Coordination Centre. (2022). *General subdirectorato for forest policy and the fight against desertification*. <http://www.miteco.gob.es>
- Pavía, J. M., & Aybar, C. (2018). Field rules and bias in random surveys with quota samples. An assessment of CIS surveys. *Statistics and Operations Research Transactions*, 42(2), 183–206. <https://doi.org/10.2436/20.8080.02.74>
- Pérez, V., Aybar, C., & Pavía, J. M. (2022). Dataset of the COVID-19 post-lockdown survey conducted by GIPEyOP in Spain. *Data in Brief*, 40, Article 107763. <https://doi.org/10.1016/j.dib.2021.107763>
- Ramalho, A. J. (2018). Relaciones cívico-militares en Brasil. Los retos del sector defensa y el rol de los militares en la sociedad brasileña contemporánea [Civil-military relations in Brazil. The challenges of the defense sector and the role of the military in contemporary Brazilian society]. In S. Namihas (Ed.), *El nuevo rol de las fuerzas armadas* [The new role of the armed forces] (pp. 11–44). Biblioteca Nacional del Perú.
- Roldán Pascual, J. E. (2012). UME. *Revista Ejército*, 857, 40–43.
- Takeshita, S., Toda, H., Tanaka, T., Koga, M., Yoshino, A., & Sawamura, T. (2020). Psychological and physical condition of Japan maritime self-defense force personnel who performed disaster-relief missions after the 2011 great East Japan earthquake. *Journal of Psychiatric Research*, 130, 104–111. <https://doi.org/10.1016/j.jpsychires.2020.07.023>
- Tarilonte, E. (2020). Misión Baluarte: una defensa contra el COVID-19 [Baluarte Mission: a defense against COVID-19]. *Revista Española de Defensa*, 376 6–13.
- Tur-Prats, A., & Valencia Caicedo, F. (2020). The long shadow of the Spanish civil war. *SSRN Electronic Journal* Advance online publication. <https://doi.org/10.2139/ssrn.3659549>

Author Biographies

Virgilio Pérez, MSc in economics, is a PhD candidate at the University of Valencia, Spain, and member in the economics department of the University of Valencia of the Research Unit of Defence, Cybersecurity and Protection. He currently works in the area of Information Systems at the Ministry of Defence and as adjunct professor in quantitative methods at the Faculty of Economics of the University of Valencia.

Cristina Aybar, MSc in economic and business sciences and PhD in economics, is associate professor in quantitative methods at the University of Valencia, Spain. She is director in the applied economics department of the University of Valencia of the Research Unit for Defence, Cybersecurity and Protection.

Jose M. Pavía, MSc in maths and PhD in economics, is quantitative methods professor in the University of Valencia, Spain. He is director of the Elections and Public Opinion Research Group of the University of Valencia, and chair of Catedra Deblanc for the application of statistical, economic, and machine learning to detect money laundering and financial crimes.

Anexo III

Escalas de Likert, auto-ubicación ideológica y encuestas del CIS

Pérez, V., Aybar, C. & Pavía, J. M. (2022). *En revisión*.

Escalas de Likert, auto-ubicación ideológica y encuestas del CIS

Resumen. Desde 1976, año en el que el CIS definió la auto-ubicación de la escala ideológica de 1 a 10, hasta el momento actual, en el que ha puesto a disposición con ficheros de microdatos, más de 1400 estudios, la ideología ha sido medida con hasta 8 escalas numéricas diferentes: 1-3; 1-4; 1-5; 1-6; 1-7; 0-10; 1-10 (la más utilizada); y una escala con opciones definidas por pares de valores. Escalas numéricas que presentan distintas variantes dependiendo de los vocablos que las acompañan, tales como “extrema izquierda”, “extrema derecha” (o simplemente “izquierda”, “derecha”) y ocasionalmente también “centro-izquierda”, “centro”, “centro-derecha” con la inclusión o no otras opciones: “apolítico”, “depende”, “no acepta esta distinción”, “ninguna”. En total identificamos 28 escalas diferentes. Esta investigación estudia, desde una perspectiva empírica, las estructuras de relaciones entre las distintas escalas y analiza y cuantifica su impacto en la estimación de la distribución ideológica de la población española.

Palabras clave: Cuestionarios, encuestas, constructos subyacentes, Centro de Investigaciones Sociológicas; España.

Abstract. From 1976, when the CIS defined the self-location ideological scale from 1 to 10, until the present moment, in which it has made available with microdata files, more than 1,400 studies, ideology has been measured using up to 8 different numerical scales: 1-3; 1-4; 1-5; 1-6; 1-7; 0-10; 1-10 (the most used); viii) a scale with options defined by pairs of values. Numerical scales that present different variants depending on the words that accompany them, such as “extreme left”, “extreme right” (or simply “left”, “right”) and occasionally also “center-left”, “center”, “center- right” with the inclusion or not of other options: “apolitical”, “depends”, “does not accept this distinction”, “none”. In total we identified 28 different scales. This research studies, from an empirical perspective, the relationship structures between the different scales and analyses and quantifies their impact on the estimation of the ideological distribution of the Spanish population.

Keywords: Questionnaires, surveys, underlying constructs, Center for Sociological Research; Spain.

Introducción

Conocer el posicionamiento ideológico de los electores es relevante para diseñar estrategias electorales, definir políticas y determinar el contenido de los programas sociales y económicos de partidos políticos y candidatos. Medir la ideología política, sin embargo, no es tarea fácil pues no depende únicamente de una dimensión. Preguntas acerca del sentimiento nacionalista del encuestado, su perfil conservador, liberal o progresista, o el recuerdo de voto, son variables que pueden ayudar a delimitar, no sin errores, un perfil ideológico (Sánchez y Sánchez 2009; Bauer et al. 2017; Sánchez 2017; Zuell y Scholz 2019; Carroll y Kubo 2021; Sinclair 2022). En realidad, preguntar sobre la auto-ubicación ideológica, escalando las posiciones de izquierda a derecha, en sí ya condiciona a una polarización en la respuesta, motivada, en gran parte, por los mismos partidos políticos (Cano 2015; Lelkes 2016; Todósijevic 2016; Gestefeld et al. 2022).

Pese a todas las dimensiones que pueden configurar la ideología de una persona, la ideología valorada mediante una graduación numérica, definida/interpretada por el propio encuestado, representa una escala ampliamente utilizada para informar sobre las tendencias políticas agregadas de un país, región o comunidad. De hecho, tal y como en algún momento indicaba el Centro de Investigaciones Sociológicas (CIS) en su propia pregunta de auto-ubicación ideológica, escalar en una serie de números dicha ubicación, interpretándola de izquierda a derecha, se puede observar como un termómetro político de la sociedad (ver, por ejemplo, imagen 24 en la Sección S1 del material suplementario). Se ha de notar, no obstante, que el término “termómetro” se ha dejado de utilizar ya que no se comprendía bien si izquierda significaba frialdad y derecha calentura y si ello pudiera estar induciendo a posiciones centrales “más saludables”.

Si bien la confiabilidad de las escalas ha sido una de las cuestiones que más interés ha despertado entre los investigadores de las escalas de Likert, nuestra investigación adopta una aproximación algo diferente. Nuestro trabajo se centra, en el contexto de las encuestas del CIS y de la medición de la ideología, en estudiar qué implica (cuáles son las consecuencias de) emplear distintas escalas de medida. Distintas escalas que varían no solo por el número de categorías de respuesta, sino también por el etiquetado de las mismas.

Este documento se organiza como sigue. Tras esta introducción, en la sección segunda, se fijan los objetivos y se realiza una breve revisión de antecedentes. La sección 3 presenta una descripción de los estudios del CIS centrada en la auto-ubicación ideológica: clasificamos en grandes bloques temáticos los casi 2000 estudios analizados, determinamos en cuántos de ellos se mide la variable de auto-ubicación ideológica y si se dispone o no de microdatos, detallamos las 28 escalas distintas utilizadas por el CIS y mostramos cómo su uso ha ido cambiando a lo largo del tiempo. En el punto

4, tratamos los 1937 estudios que disponen de la variable ideología para presentar la evolución del promedio anual (1972-2022) de la auto-ubicación ideológica en escala 1-10. La sección quinta explora el impacto de la escala utilizada comparando las distribuciones de ideología agregadas que se obtienen para los mismos momentos del tiempo con escalas diferentes. La sección sexta profundiza en el análisis focalizándose en los cinco estudios que preguntaron a los mismos encuestados su auto-ubicación ideológica utilizando dos escalas diferentes. Mostramos para cada uno de estos estudios las relaciones de transferencia entre escalas. Finalizamos con una discusión de los resultados y las conclusiones. Este trabajo se completa con un material suplementario donde se amplían los detalles de esta investigación.

Objetivos y antecedentes

Existen numerosas escalas numéricas para definir la auto-ubicación ideológica. Circunscribiéndonos exclusivamente a la propia experiencia del CIS, estas varían desde una escala 1 a 3 a una escala 0 a 10, incluyendo un gran abanico de opciones intermedias. De hecho, desde que en 1932 Rensis Likert publicase su trabajo seminal sobre la medición de constructos subyacentes (Likert 1932), el estudio del número de puntos/posiciones que debe contener una escala de Likert ha sido un tema ampliamente investigado en la literatura desde una multitud de disciplinas y empleando diferentes enfoques y perspectivas. El paso del tiempo no ha mitigado el atractivo de estos temas, pues en la actualidad continúan atrayendo el interés de los investigadores (Kroh 2007; Willits et al. 2016; Matas 2018; Taherdoost 2019; Aybek y Toramen 2022). Algunas investigaciones recientes donde se puede encontrar una extensa revisión de la literatura que cubre una gran parte de los aspectos relevantes que inciden en una escala de Likert incluye los trabajos de Joshi et al. (2015), Jebb et al. (2021) y Alabi y Jelili (2022).

La enorme variedad de escalas utilizadas impele a formularse una gran cantidad de preguntas, tales como si el hecho de utilizar un número de puntos u otro en una escala ¿puede condicionar el resultado?, o sobre cuáles son las diferencias entre escalar de 1 a 10 o de 0 a 10, ¿es posible que los encuestados no se identifiquen con el 0 porque lo vean como un número con connotaciones negativas?, o ¿es posible que se elija en ambas escalas el 5 como centro, pese a que la primera escala no tiene un centro específico definido explícitamente? Otras posibles preguntas incluyen ¿por qué es más habitual el uso de la escala 1-10? (Dassonneville 2020). La escala 1-10 no tiene un centro sino dos, los valores 5 y 6, aunque habitualmente se selecciona el valor 5, sesgando las respuestas hacia la izquierda (del Castillo 2004), ¿es esto idóneo? ¿Cuáles son sus implicaciones? ¿Por qué no emplear una escala 1-9 con un centro claro y simetría en los ejes izquierda-derecha? (Nilsson y Erlandsson

2015). La conveniencia de utilizar una u otra escala ¿dependerá del objeto de estudio? o, como indican Bisquerra y Pérez-Escoda (2015), la escala de 11 puntos (0-10) es la idónea.

El análisis es todavía más complejo pues además de valores numéricos, a veces las escalas también incluyen aclaraciones, generalmente parciales, con etiquetas de texto. Rótulos que se suelen ubicar habitualmente en los dos puntos extremos (o anclas) de la escala y también, por momentos, en otras posiciones. Las cuestiones adicionales que surgen son, por tanto, determinar cuáles son los impactos, si existen, de estas prácticas.

En efecto, junto con números en los puntos intermedios, habitualmente solo los extremos de la escala de auto-ubicación ideológica contienen etiquetas, como “izquierda” y “derecha” o “extrema izquierda” y “extrema derecha”, y ello a pesar de la existencia de investigaciones (e.g., Krosnick 1991) que abogan por la conveniencia de utilizar escalas completamente etiquetadas para una mejor comprensión de la pregunta por parte de los encuestados. La literatura, sin embargo, no ha alcanzado consenso en este punto. Otros autores sugieren que dejar un número mayor de puntos libres, sin etiquetar, puede permitir al encuestado seleccionar una posición más cercana a su ideología política (McRoy y Pancratz 2020).

Una escala más amplia, sin embargo, genera diversos problemas prácticos y puede no representar una diferencia significativa. Por una parte, más posibilidades de elección puede plantear dificultades cognitivas de elección (Schwartz 2005) y colisiona con la recomendación de etiquetar cada uno de los valores de la escala, por la necesidad de disponer de un diccionario de vocablos cada vez mayor y por la dificultad de garantizar que las distancias entre dos vocablos contiguos sean las mismas (Cummins y Gullone 2000). Por otro lado, existen estudios que muestran una escasa sensibilidad de los resultados a cambios en el número de puntos en la escala (Evans 2019; McRoy y Pancratz 2020), e incluso hay investigaciones que analizan si la dirección de las escalas influye en las respuestas (Stefkovics 2022). Es decir, estudian el impacto de revertir la escala: si los resultados cambiarían si la escala se direcciona al revés. De hecho, aunque el CIS habitualmente utiliza una escala izquierda-derecha, en catorce de sus estudios ensayó un cambio de dirección (ver imágenes 1 a 6, 11 y 15 de la Sección S1 del material suplementario), aunque en todos esos casos las escalas estaban completamente etiquetadas.

Este trabajo busca dar respuesta a las cuestiones anteriores para el caso de la medición de la distribución de auto-ubicación ideológica de los españoles obtenida a través de las encuestas del CIS. Por una parte, mostramos (en una escala homogeneizada) la evolución temporal de la media de la escala; por otro lado, al disponer de una enorme cantidad de estudios coincidentes en el tiempo que han empleado escalas distintas estudiamos las distribuciones de auto-ubicación ideológica como

función de la escala utilizada. Es más, al disponer de cinco estudios en los que se auto-ubicaron ideológicamente las mismas personas en dos escalas alternativas, explotando las respuestas individuales, realizamos varios análisis de transferencia/relación entre escalas. Respecto a esta última cuestión se ha de notar las limitaciones que para el cálculo de estas transferencias puede suponer la imposibilidad de valorar si un cambio del orden de las preguntas hubiera producido transferencias distintas y cómo puede influir que el encuestado perciba que se le está preguntando dos veces lo mismo.

Nuestro análisis además no elude el estudio del efecto de la no respuesta, si no que le dedica especial atención. Todas las preguntas incluyen las opciones de no respuesta “no sabe” y “no contesta” y consideramos su análisis fundamental para, entre otras cosas, detectar qué posición se oculta detrás de esas respuestas. Es decir, si la elección de esas opciones de respuesta esconde una posición apolítica, un desconocimiento de su propia ideología, una incomprensión de la pregunta, una disconformidad con la escala propuesta o una abstención electoral (Camino 2022). La comparación de los estudios que sí incluyen las opciones “ninguna” o “apolítico” o “depende”, con las que no incluyen ninguna de estas opciones puede arrojar luz sobre esos silencios.

Finalmente señalar que, a pesar de haber identificado, como se muestra en la siguiente sección, 28 escalas diferentes, no hemos considerado oportuno contemplar como escalas distintas aquellas en las que la misma escala se dibuja con las opciones de respuesta en forma horizontal o vertical. Y ello a pesar de que hay estudios que comprueban que el formato de la pregunta puede condicionar la respuesta al aparentar visualmente cambios en las distancias (Weijters et al. 2021). Consideramos que, en este caso, la orientación de la escala no es tan importante como en escalas del tipo “totalmente de acuerdo”, “de acuerdo”, “ni de acuerdo ni desacuerdo”, “en desacuerdo”, “totalmente en desacuerdo”.

Descripción de los estudios del CIS respecto a la auto-ubicación ideológica

El Centro de Investigaciones Sociológicas (CIS) tiene como objetivo el estudio de la sociedad española mediante la elaboración de encuestas periódicas realizadas por propia iniciativa o a petición de otros organismos. Desde enero de 2009 el CIS difunde gratuitamente, a través de su página web (www.cis.es), los archivos de microdatos disponibles de las encuestas realizadas por el Centro junto con los metadatos necesarios para su utilización. Constantemente el CIS trata de completar la información, recuperando los microdatos de estudios para los cuales en la web solo se dispone de cuestionarios u otros archivos.

Clasificación de los estudios del CIS

A fecha de 21 de julio de 2022, en el banco de datos del CIS se dispone de información de 2176 estudios (sin considerar los estudios de carácter cualitativo). De ellos, 631 no disponen de microdatos (ver Tabla 1). Aunque, esa cifra se irá reduciendo progresivamente, ya que como hemos señalado el CIS está incorporando paulatinamente esta información faltante en su banco de datos. Los estudios que realiza el CIS cubren diversas temáticas y pueden ser agrupados atendiendo a diversas clasificaciones. En este estudio, proponemos una agrupación en cinco bloques que consideramos útil a la hora de analizar la variable ideología política, objeto de estudio en este artículo. En concreto:

- *Barómetros (BAR)*: este grupo está compuesto por un conjunto de encuestas que se realizan con periodicidad mensual (excepto los meses de agosto) y que tienen como principal objetivo medir el estado de la opinión pública española del momento. Para ello se entrevista en torno a 2.500 personas (3.000 desde septiembre de 2018), elegidas al azar, dentro del territorio nacional, de las que, además de sus opiniones, se recoge una amplia información social y demográfica para el análisis. Tradicionalmente estas encuestas se realizaban en los hogares mediante un complejo procedimiento de selección muestral (ver, e.g., Pavía y Aybar 2018), aunque desde el inicio de la pandemia de COVID-19 han pasado a ser telefónicas.
- *Índice de Confianza del Consumidor (ICC)*: este grupo está constituido por un conjunto de estudios mensuales diseñados para aproximarse a las intenciones de gasto de los consumidores, quienes son inquiridos acerca de su percepción actual y sus expectativas de futuro para la economía del país, su economía familiar y el empleo. La encuesta se realiza telefónicamente a una muestra de 3.000 individuos, mayores de 16 años, representativos del conjunto de la sociedad española.
- *Encuestas electorales (ELE)*: este grupo está formado por encuestas realizadas, en periodo electoral (pre-electoral) y/o posterior (post-electoral), sobre elecciones municipales, autonómicas, nacionales y europeas. Al igual que los barómetros han pasado a implementarse telefónicamente, asistidas por ordenador, ya que los encuestadores del CIS teletrabajan (lo mismo para los dos siguientes ítems).
- *Estudios sobre temas políticos (POL)*: este grupo está formado por encuestas no periódicas de carácter político-institucional, que no se pueden incluir en la categoría anterior, que abarcan cuestiones de Estado, Constitución e instituciones, cultura política, partidos y líderes políticos.
- *Otros estudios (EOT)*: el último grupo lo forma un conjunto de estudios que no pueden ser incluidos en los grupos anteriores y que tienen como objetivo profundizar en multitud de temas variados, como economía, trabajo, relacionados con la población (inmigración, familia

y formas de convivencia, natalidad, infancia, juventud, vejez, mujer), medio ambiente, comunicación, cultura, ocio y deporte, ciencia y tecnología, educación, sanidad, justicia, etc.

Tabla 1. Distribución de estudios con y sin microdatos e ideología.

Tipo estudio		Con ideología	Sin ideología	Total
BAR	Con microdatos	397	4	401
	Sin microdatos	93	0	93
	Total	490	4	494
ICC	Con microdatos	128	0	128
	Sin microdatos	0	1	1
	Total	128	1	129
ELE	Con microdatos	374	5	379
	Sin microdatos	204	43	247
	Total	578	48	626
POL	Con microdatos	162	4	166
	Sin microdatos	82	13	95
	Total	244	17	261
EOT	Con microdatos	389	86	475
	Sin microdatos	112	83	195
	Total	501	169	670
TOTAL	Con microdatos	1446	99	1545
	Sin microdatos	491	140	631
	Total	1937	239	2176

Nota: El número de estudios clasificados en cada combinación se ha obtenido a partir de la información disponible a 21 de julio de 2022. El estudio 2743, clasificado en esta investigación con el prefijo POL, ha sido desglosado en dos (2743A y 2743B) y tratado como dos estudios independientes, al aplicarse a dos muestras distintas. Los estudios 3242, 3245, 3263 y 3269 han sido incluidos a efectos de esta estadística en dos grupos (BAR y ELE) al corresponder a ambas agrupaciones.

Para comprender la importancia que la medición la auto-ubicación ideológica tiene para el CIS, en la Tabla 1 se ofrece la distribución del número de estudios agrupados según la clasificación anterior detallando el número de estudios que indagan sobre esta cuestión (auto-ubicación ideológica). La tabla también ofrece el número de estudios de cada tipo para los cuales se dispone o no de microdatos a 21 de julio de 2022. Por una parte, la Tabla 1 muestra que, de un total de 2176 encuestas, 1937 estudios incluyen la variable de interés; un 89,02%. Por otra parte, la tabla también refleja el gran esfuerzo que viene realizando el CIS recuperando estudios pasados y ofreciendo los microdatos en su web. A 21 de julio de 2022, se observa que el 71% de ellos tenían microdatos disponibles.

La categoría con más estudios es ELE, que engloba las encuestas electorales. El número total de estudios electorales puede hacer pensar que para medir la auto-ubicación ideológica del conjunto de los españoles este es el grupo más numeroso, pero este dato puede resultar engañoso ya que la mayoría de ellos no cubren el total de la población española. En elecciones autonómicas y municipales, por

ejemplo, se dispone para cada elección de un estudio por autonomía y para los principales municipios. Además, es en este conjunto donde restan más estudios por rescatar los microdatos (no están disponibles en el 39,46% de ellos). Tanto los barómetros como los indicadores de confianza del consumidor son los estudios que, además de ser los más regulares en el tiempo, cuentan con una mayor información acerca de la variable de escala ideológica.

Taxonomía de escalas empleadas por el CIS

A lo largo de los años (desde 1976 a 2022), la pregunta sobre la auto-ubicación ideológica se ha planteado por el CIS con 8 escalas numéricas diferentes que, combinadas con las diferentes etiquetas de texto empleadas y las distintas opciones usadas para capturar otras alternativas de respuesta distintas a declarar una auto-ubicación, han dado lugar a 28 formas distintas de medición. Estas 28 escalas difieren en el número de puntos, en el texto que informa sobre la definición de las escalas (es decir, en cómo se etiquetan el número inferior, superior o los valores centrales de la escala) y/o en la inclusión o no de la opción “Ninguna” y/o “Apolítica” o “Depende”. De hecho, dos escalas identificadas como distintas pueden compartir el mismo enunciado de la pregunta y gráficamente estar representadas igual, pero diferir en otros aspectos como la inclusión de opciones de respuesta fuera de la escala.

Durante la construcción de la taxonomía de escalas hemos mantenido el orden cronológico de su uso por el CIS. En la sección S1 del material suplementario se muestra una imagen de cada una de las mediciones, donde se ha utilizado como imagen representante de cada escala la correspondiente al primer estudio en el que se usó dicha escala. En concreto, la agrupación en 8 escalas numéricas, con la especificación concreta asociada a las 28 escalas en que se desdoblan, se muestra a continuación:

Escala 1-3:

(E1) Derecha, centro, izquierda.

Escala 1-4:

(E2) Derecha, centro-derecha, centro-izquierda, izquierda.

Escala 1-5:

(E3) Derecha, centro-derecha, centro, centro-izquierda, izquierda.

(E4) Extrema derecha, derecha, centro, izquierda, extrema izquierda, apolítico.

(E5) Extrema derecha, derecha, centro, izquierda, extrema izquierda.

(E6) Extrema derecha, derecha, centro, izquierda, extrema izquierda, depende.

(E7) Izquierda, centro-izquierda, centro, centro-derecha, derechas.

(E8) Extrema izquierda, izquierda, centro, derecha, extrema derecha.

Escala 1-6:

(E9) Extrema izquierda, izquierda, centro-izquierda, centro-derecha, derecha, extrema derecha.

Escala 1-7:

(E10) Extrema izquierda, extrema derecha.

(E11) Extrema derecha, derecha, centro-derecha, centro, centro-izquierda, izquierda, extrema izquierda, no acepta esta distinción.

(E12) Extrema izquierda, izquierda, centro-izquierda, centro, centro-derecha, derecha, extrema derecha.

(E13) Extrema izquierda, izquierda, centro-izquierda, centro, centro-derecha, derecha, extrema derecha, ninguna, apolítico.

(E14) Izquierda, derecha.

(E15) Extrema derecha, derechas, centro-derecha, centro, centro-izquierda, izquierdas, extrema izquierda.

(E16) Extrema izquierda, izquierda, centro-izquierda, centro, centro-derecha, derecha, extrema derecha, ninguna.

Escala 1-10:

(E17) Izquierda, derecha.

(E18) Extrema izquierda, extrema derecha.

(E19) Sin texto.

(E20) Extrema izquierda, izquierda, centro, derecha, extrema derecha.

(E21) Más a la izquierda, más a la derecha.

(E22) Izquierda, centro, derecha.

(E23) Izquierda, derecha, ninguna.

(E24) Lo más a la izquierda, lo más a la derecha.

Escala 0-10:

(E25) Extrema izquierda, extrema derecha.

(E26) Izquierda, derecha, ninguna.

(E27) Izquierda, derecha.

Escala por pares:

(E28) En la ficha del cuestionario (estudio 2692) aparece la escala 1 a 10, pero en los microdatos se recogen los pares (1-2), (3-4), (5-6), (7-8), (9-10), motivo por el que en la sección S1 del material suplementario no se muestra imagen asociada a esta escala.

En el resto del documento, identificaremos cada escala con el código entre paréntesis de la lista anterior.

Tabla 2. Distribución de estudios según tipo de escala

escala	número de estudios		escala	número de estudios	
	sin microdatos	con microdatos		sin microdatos	con microdatos
E1	1	0	E15	0	5
E2	1	0	E16	0	1
E3	6	0	E17	288	1258
E4	1	0	E18	55	41
E5	1	0	E19	1	1
E6	1	0	E20	0	1
E7	1	0	E21	0	1
E8	0	2	E22	0	2
E9	2	1	E23	0	15
E10	125	31	E24	0	44
E11	0	1	E25	1	2
E12	2	2	E26	0	20
E13	2	0	E27	1	21
E14	2	1	E28	0	1

Nota: Cinco estudios contienen una doble escala de la auto-ubicación ideológica: POL1163 (E10 y E11); OTE2731 (E9 y E25); POL2743A (E8 y E17); POL2743B (E8 y E27); ELE3198 (E16 y E17).

El hecho de que se hayan identificado 8 grupos y 28 escalas distintas, no implica que todas ellas sean igualmente frecuentes, muy al contrario. El número de estudios que hay en cada una de las escalas, distinguiendo para los que se dispone o no de microdatos, se presenta en la Tabla 2. Los datos de la tabla muestran que en el 80% de los estudios en los que se pregunta la auto-ubicación ideológica la escala utilizada es E17 (1-10), identificando 1 con “Izquierda” y 10 con “Derecha”. Si nos fijamos únicamente en los estudios con microdatos disponibles, ese porcentaje aumenta hasta el 87%. El segundo grupo más numeroso, a gran distancia al ser diez veces menos frecuente, está integrado por las encuestas que utilizan la escala E10 (1-7), donde al 1 se le añade la etiqueta “Extrema izquierda” y al 7 “Extrema derecha”. En todo caso, las escalas integradas por 10 y 7 niveles/puntos han sido las más empleadas, un 96,7% de las ocasiones.

El detalle de relación entre escala y estudios se ofrece en la sección S2 del material suplementario. Allí, se incluyen dos tablas, una para los estudios donde no se dispone de microdatos y otra para los

estudios donde estos sí están disponibles, con la identificación de los estudios que tienen la variable escala ideológica según cada una de las 28 tipologías listadas anteriormente. Los estudios son identificados utilizando un prefijo que identifica la tipología de estudio (BAR, barómetros; ICC, indicadores de confianza del consumidor; ELE, estudios electorales; POL, estudios de política; y EOT, otros temas) seguido por 4 dígitos numéricos que informan sobre el número a que corresponde el estudio en el banco de datos del CIS.

Análisis temporal del uso de las escalas por parte del CIS

En la Figura 1 se muestra, en perspectiva temporal, la frecuencia con que cada una de las escalas han sido utilizadas a lo largo del tiempo en los cuestionarios del CIS. Como se puede apreciar, hasta finales de 1983 la escala más utilizada fue la E10 (1-7), identificando 1 con “Extrema izquierda” y 7 con “Extrema derecha”, siendo a partir de 1983 cuando se comenzaron a utilizar con mayor asiduidad las escalas 1-10: de 1983 a 1985 la escala E18, identificando 1 con “Extrema izquierda” y 10 con “Extrema derecha”, y a partir de 1985 la escala E17. A finales de 2019 se comenzó a utilizar con frecuencia la escala E24 (1-10), identificando el 1 con “Lo más a la izquierda” y el 10 con “Lo más a la derecha”, evitando así el uso de la palabra “Extrema”.

Aunque como se observa claramente en la Tabla 1 y en la Figura 1, la escala E17 (1-10) ha sido la más utilizada (especialmente después de 1985), su uso no es homogéneo entre categorías de estudios. Dependiendo de la tipología del estudio (BAR, ICC, ELE, POL y EOT), podemos encontrar diferencias importantes en el uso de las escalas. La Figura 2 ofrece el detalle de cómo ha ido cambiando en cada grupo de estudios el uso de las diferentes escalas a lo largo del tiempo. Los cambios más destacados se observan en el grupo de barómetros y en el de otros estudios.

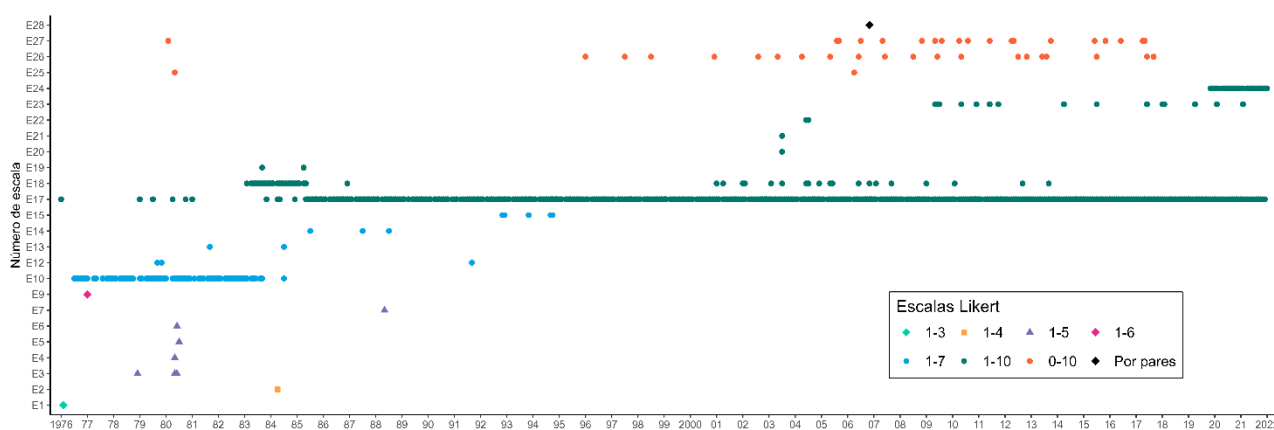


Figura 1. Evolución temporal en el uso de las diferentes escalas por el CIS (1976-2022).

En los 490 barómetros (BAR) disponibles que miden la variable ideología, el CIS solo ha ensayado, del total de 28 escalas distintas identificadas en este trabajo, cinco escalas diferentes: la escala E10

(en 39 estudios), E17 (396 estudios), E18 (29 estudios), E19 (1 estudio) y E24 (25 estudios). Desde un punto de vista temporal, se aprecian varios fenómenos. En los primeros barómetros (de 1979 a 1983) se utilizaba la escala E10 (1-7). A partir de 1983 se comenzó a utilizar la escala E18 (1-10) que incluye el término “Extrema”, pasando en 1985 a utilizarse la escala E17 (1-10), haciendo referencia únicamente a “Izquierda/Derecha”. A finales de 2019 se produce un nuevo cambio, la escala E17 (la más utilizada tanto en barómetros como en el resto de tipos de estudio) es reemplazada por la E24 (1-10), que utiliza las expresiones “Lo más a la ...” izquierda o derecha.

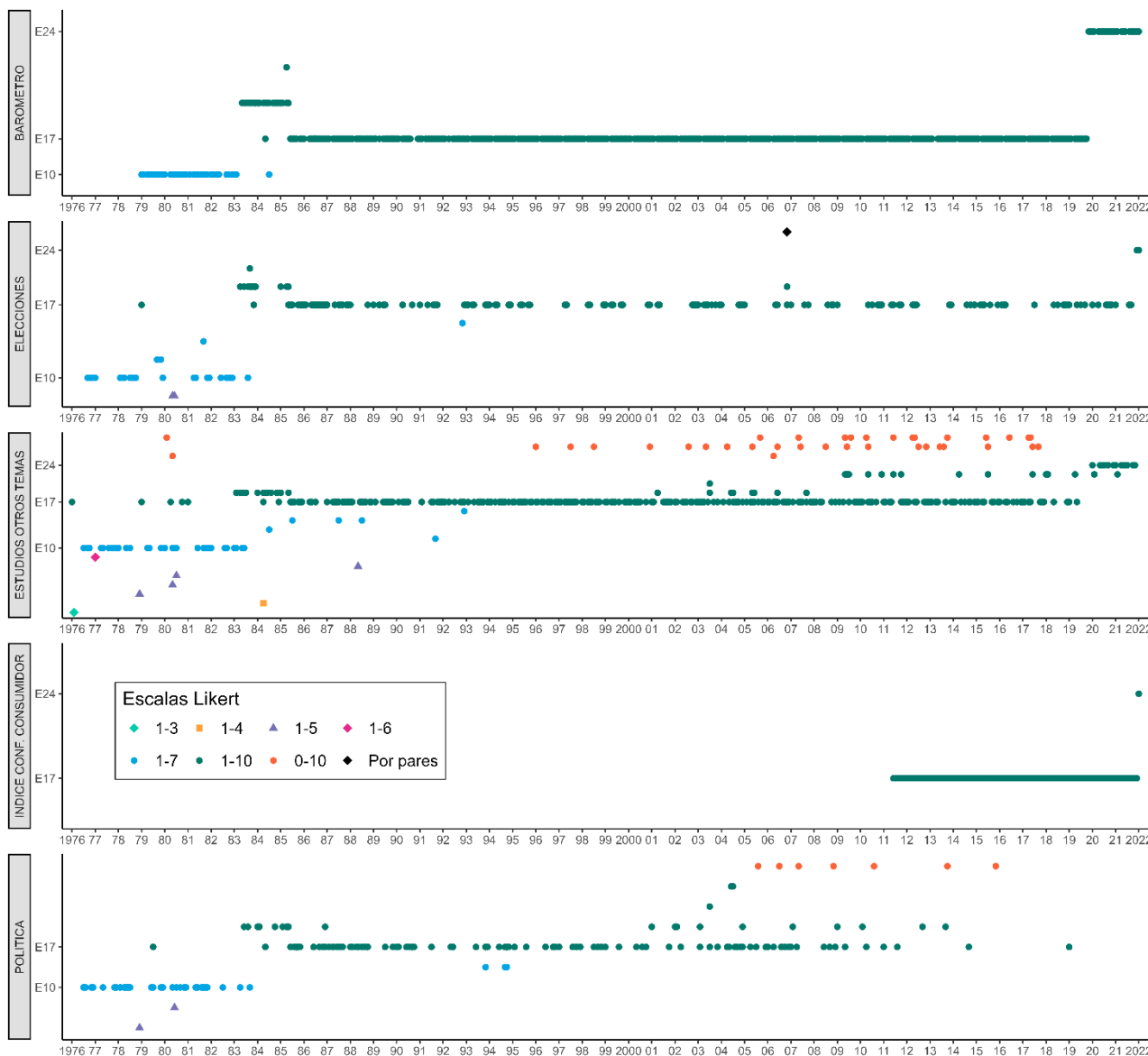


Figura 2. Evolución temporal en el uso de las diferentes escalas utilizadas por el CIS (1976-2022) diferenciando tipo de estudio.

En el grupo de encuestas ELE ocurre algo similar a lo que ocurre dentro del grupo BAR, especialmente en los primeros años, si bien a lo largo del tiempo en este tipo de estudios se ha ensayado un abanico de escalas de medición más amplio: la escala E3 (en 4 estudios), E10 (33

estudios), E12 (3 estudios), E13 (1 estudio), E15 (1 estudio), E17 (514 estudios), E18 (17 estudios), E19 (1 estudio), E24 (2 estudios) y E28 (1 estudio).

Dentro del grupo de otros estudios EOT es donde encontramos encuestas más escalas utilizadas distintas: 20. De nuevo, no obstante, es la escala E17 la que ha sido con diferencia más utilizada: en 348 estudios. Respecto al resto de escalas, las escalas E25-E27 (0-10) han sido utilizadas en estudios de este grupo hasta en 37 ocasiones, lo que representa el 82% de las 45 veces que las escalas 0-10 han sido empleadas por el CIS. Observando más en detalle cuando estas escalas han sido empleadas, encontramos ciertos patrones en cuanto a temáticas. En concreto, se ha utilizado en 15 “Latinobarómetros” y 8 estudios sobre “Actitudes hacia la inmigración”. Los 14 estudios restantes tienen una temática más heterogénea (cultura democrática, ocio nocturno, atribución de responsabilidades, etc.). Desde una perspectiva temporal, el uso de las escalas en el grupo EOT se asemeja mucho a la de los grupos BAR y ELE, caracterizados por: una primera etapa donde se empleó mayoritariamente la escala E10, un pequeño periodo de transición (1983-1985) donde se mayoritariamente usó la escala E18, una segunda etapa (de casi 35 años) en que sistemáticamente se usó la escala E17 y el periodo actual en el que la escala de referencia es la E24. De hecho, la sensación es que el CIS está realizando una transición hacia emplear la escala E24 en (casi) todos sus estudios.

La posible validez de la última observación anterior viene reforzada por el hecho de que en las encuestas del grupo ICC de manera sistemática siempre se había venido utilizado la escala E17, produciéndose un cambio en el último estudio de este tipo recogido en esta investigación. En junio de 2022 el ICC abandonó la escala E17 y pasó a medir la auto-ubicación ideológica utilizando la escala E24.

Por último, analizando los estudios agrupados bajo el epígrafe POL, encontramos que en este grupo se han utilizado hasta 9 escalas de medición diferentes desde 1976. Durante los primeros años se utilizó con asiduidad la escala E10 (en 44 estudios), hasta que en 1984 se comenzaron a utilizar las escalas E17 y E18 (163 y 19 estudios, respectivamente). En 2006 se comenzó a utilizar también la escala E27 (7 estudios). El resto de escalas utilizadas son: E3 (1 estudio), E6 (1 estudio), E15 (3 estudios), E20 (1 estudio) y E22 (2 estudios).

Evolución temporal de la auto-ubicación ideológica media

En las escalas de Likert, las categorías de respuesta se disponen de manera que las distancias entre categorías contiguas sean lo más homogéneas posible, pudiendo contener un punto neutral (DeVellis 2003), con esta posición neutral correspondiendo a la ideología política de centro en el contexto político. La cuestión de si “deben procesarse los datos en una escala ordinal o en una escala de

intervalos iguales” se plantea a menudo. De acuerdo con Norman (2010), las escalas de Likert son asimilables a escalas de intervalos iguales y pueden ser usadas en análisis paramétricos con esta suposición. De hecho, estudios basados en simulación (Harpe 2015; Huiping y Shing-On 2017; Mircioiun y Atkinson 2017; Wu y Leung 2017; Canto de Gante et al. 2020) muestran que aumentar el número de calificaciones en la categoría de respuesta de una escala de Likert daría como resultado una distribución normal y una similitud con una escala de intervalo. El propio CIS trata la variable de auto-ubicación de la escala ideológica como cuantitativa, al utilizar su valor medio como indicador de referencia y su desviación típica como medida para verificar la validez del cálculo de la media (del Castillo 2004).

De acuerdo con los argumentos anteriores, por tanto, se considera lícito realizar transformaciones cuantitativas a los datos, tratando las observaciones como valores numéricos medidos en una escala de intervalo. El CIS ha utilizado a lo largo de los años diferentes escalas, con distinto número de puntos, lo que dificulta su comparación, por lo que, a fin de estudiar la evolución temporal media de auto-ubicación ideológica en España, todas las respuestas se han trasladado a una escala 1-10, utilizada en el 87% de los estudios con microdatos disponibles. Las escalas del 13% de estudios restantes han sido homogeneizadas (re-escaladas) a la escala 1-10 mediante:

$$x_k = x_{k-1} + \frac{9}{n-1}$$

donde $x_1 = 1$, k toma valores de 1 a 10 y n es el número de opciones de la escala inicial.

Hemos de notar, no obstante, cierta debilidad que, a nuestro juicio, presenta la escala 1-10 respecto a su valor medio percibido. El valor medio percibido por el respondiente suele ser 5 cuando el valor medio real asociado a la escala es 5,5. Esto puede provocar cierta disonancia, distorsionando los análisis y sesgando ligeramente los resultados hacia la izquierda.

Una vez homogeneizados todos los estudios disponibles, podemos analizar cómo ha evolucionado temporalmente en promedio la auto-ubicación ideológica de los españoles: ver Figura 3. Como se puede apreciar en la Figura 3, la media anual de auto-ubicación ideológica siempre es inferior al valor medio teórico de la escala, observándose también que los cambios de Gobierno siempre vienen precedidos de varios años donde la tendencia en la evolución de la media varía alejándose de la zona teórica del partido político gobernante, acercándose hacia el área del principal partido de la oposición. Antes de la llegada al poder del partido socialista en 1982, la tendencia de la media de auto-ubicación ideológica muestra un descenso hacia posiciones de izquierda, que se repite en sentido contrario hacia final de la primera etapa de gobierno socialista: desde 1989 (cuando gobierna el PSOE, partido de izquierdas) el promedio va aumentando (se sitúa cada vez más a la derecha) hasta que en 1996 entra

en el poder el PP (partido de derechas). Este ciclo parece se repite en las siguientes ocasiones que se produce un cambio de gobierno. A partir del año 2000 el promedio va disminuyendo (se sitúa cada vez más a la izquierda) hasta que en 2004 entra en el gobierno el PSOE. Asimismo, desde 2008 el promedio va aumentando (se sitúa cada vez más a la derecha) hasta que en 2011 entra en el gobierno nuevamente el PP. De nuevo, desde 2015 el promedio se sitúa cada vez más a la izquierda hasta que en 2018 entra en el gobierno nuevamente el PSOE. Finalmente, desde 2020 el promedio se sitúa cada vez más a la derecha, la pregunta, por tanto, es: ¿qué ocurrirá en las elecciones generales de 2023?

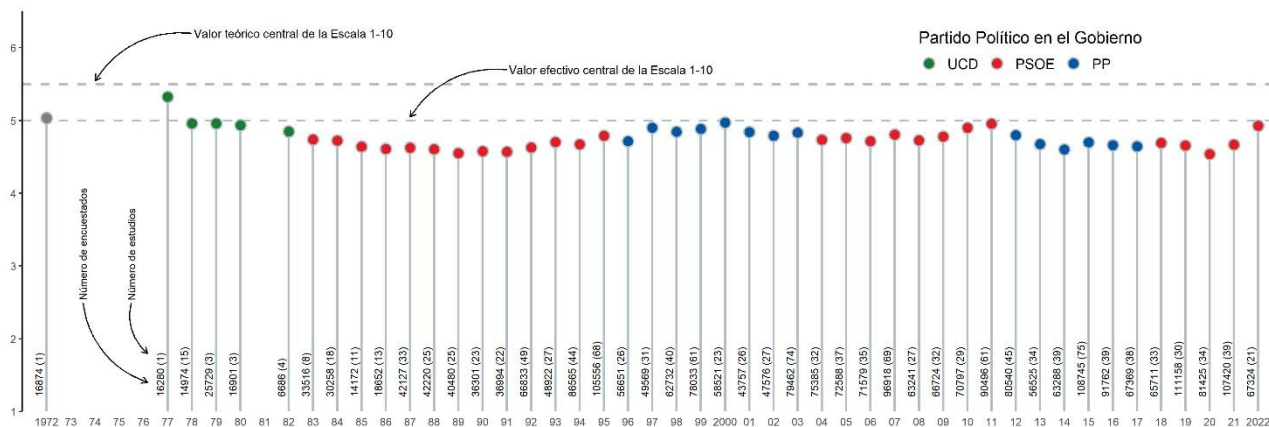


Figura 3. Evolución del promedio anual (1972–2022) de la auto-ubicación ideológica en escala 1-10. Los valores han sido re-escalados a la escala 1-10 en los estudios con una escala distinta.

Impacto de la escala en la distribución de auto-ubicación ideológica de los españoles

Cómo se ha mostrado en la sección 3, el CIS ha utilizado a lo largo del tiempo una gran variedad de escalas, con muchas de las ellas siendo utilizadas simultáneamente, en diferentes estudios cuyos trabajos de campo tuvieron lugar en los mismos momentos del tiempo. Es posible, por tanto, considerar un esquema típico de ensayos clínicos con emparejamientos caso-control y comparar las distribuciones que se obtienen midiendo, en muestras independientes, la misma realidad (la distribución de auto-ubicación ideológica de los españoles) en el mismo momento del tiempo con escalas diferentes. Para los emparejamientos solo consideramos encuestas cuyo universo coincide con el conjunto de la población española (residente en España) y cuyo periodo de recogida de datos (trabajo de campo) tuvo lugar en el mismo mes y año. Las Figuras 4 a 9 presentan las distribuciones que se derivan de los emparejamientos formados, su análisis ofrece respuestas a gran parte de las preguntas de investigación planteadas en la sección 2.

Las Figuras 4 y 5 muestran las distribuciones que se obtienen utilizando dos escalas 1-10 que solo difieren en cómo se etiquetan los extremos. La Figura 4 muestra que el uso de la palabra “extremo” no parece tener un efecto apreciable en términos de distribución izquierda/derecha: el 72,7% y el

74,2% de la población se ubican en la parte teóricamente izquierda (1-5) de la escala. Sin embargo, contrariamente a lo esperable, el uso del adjetivo “extrema” ubica un mayor número de respondientes en los límites de la escala. En el caso de los votantes de izquierda, sin embargo, actúa como repulsivo, al vaciar relativamente las puntuaciones 2 a 4 e incrementar sensiblemente el porcentaje de encuestados que se sitúan en el 5: 39,8% vs 29,1%.

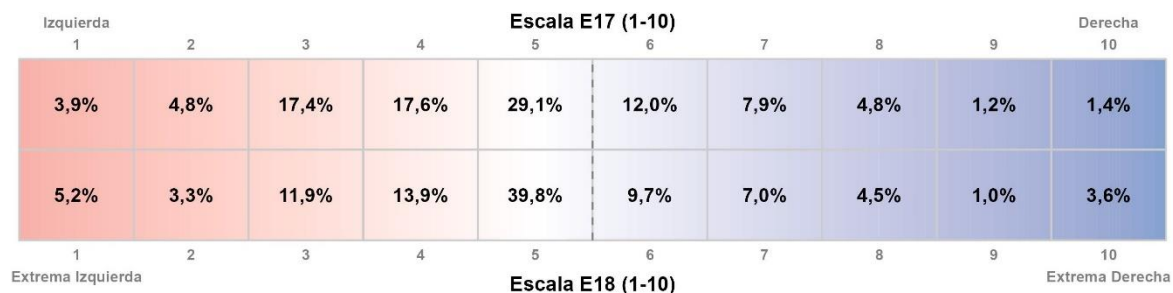


Figura 4. Distribuciones de auto-ubicación ideológica obtenidas con las escalas E17 y E18 de muestras recolectadas en los mismos momentos del tiempo. Detalles en sección S3.1 del material suplementario.

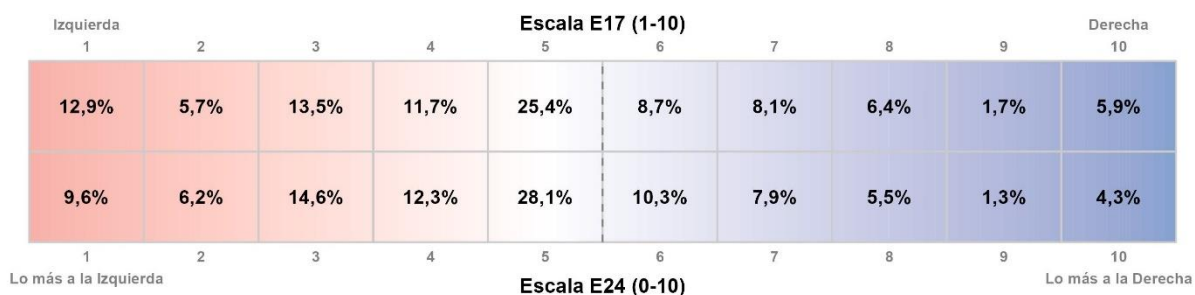


Figura 5. Distribuciones de auto-ubicación ideológica obtenidas con las escalas E17 y E24 de muestras recolectadas en los mismos momentos del tiempo. Detalles en sección S3.1 del material suplementario.

Analizando los resultados de la Figura 5, obtenemos, en términos de distribución izquierda/derecha, las mismas conclusiones: el 69,2% se auto-ubica por debajo de 5 con la escala E17 y el 70,7% con la escala E24. En términos de los extremos de la distribución, el uso de las expresiones “lo más a...” actúa como repulsor, especialmente de nuevo en la izquierda, empujando a los encuestados hacia posiciones más centradas.

La Figura 6 muestra el resumen de las distribuciones que se obtienen en los mismos momentos del tiempo con las escalas E17 y E27. La escala E27 difiere de la E17 en que tener una categoría más (cero) y un centro geométrico bien definido. La comparación entre estas dos escalas permite reforzar un resultado ya previamente identificado en la literatura (e.g., del Castillo 2004): la puntuación 5 es interpretada, por una mayoría de españoles, como centro en escalas con valor máximo 10, independientemente de si ese valor es o no efectivamente el centro geométrico. De hecho, en la escala

E17 el 43,4% de los encuestados se sitúa entre 1 y 4 (es decir, en lo que podemos considerar como izquierda si asumimos que 5 es el centro de la escala), mientras que en la escala E27 el 45,8% de los entrevistados se sitúan entre 0 y 4. También se observa que el efecto de arrancar la escala en 0 es menor, al no suponer un desincentivo para elegir el extremo inferior de la escala.

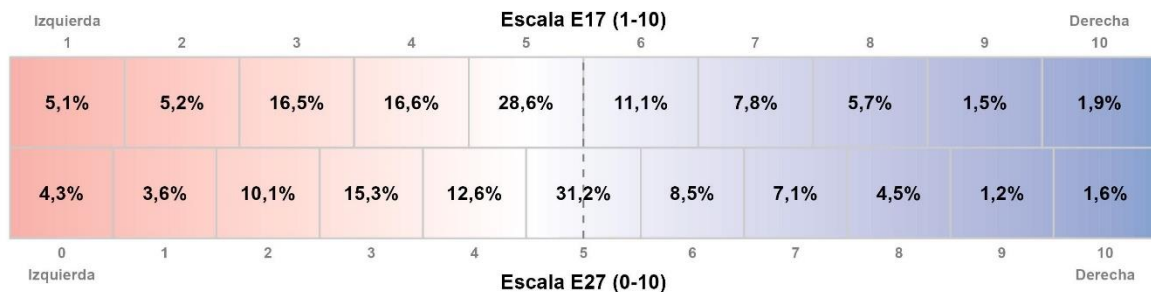


Figura 6. Distribuciones de auto-ubicación ideológica obtenidas con las escalas E17 y E27 de muestras recolectadas en los mismos momentos del tiempo. Detalles en sección S3.1 del material suplementario.

Las Figuras 7 y 8 muestran las distribuciones que se obtienen en dos escalas diferentes (con 10 niveles y 11 niveles, respectivamente), pero que difieren en la inclusión o no de una alternativa de respuesta, identificada con la etiqueta “Ninguna”, fuera de la escala. Ambos pares de distribuciones son en general bastante similares, focalizándose sus principales diferencias en la cantidad de entrevistados que se auto-ubican en el 5, identificado como centro de ambas escalas por la mayoría de los ciudadanos. El peso del 5 es ligeramente inferior en la escala que incluye la etiqueta “Ninguna”, apuntando a que una parte de los ciudadanos que se caracterizan por no tener ninguna ideología tenderían a interpretar el punto central de la escala como equivalente a la indiferencia, al no posicionamiento, prefiriendo auto-ubicarse en el centro de la escala a no responder (quizás por deseabilidad social).

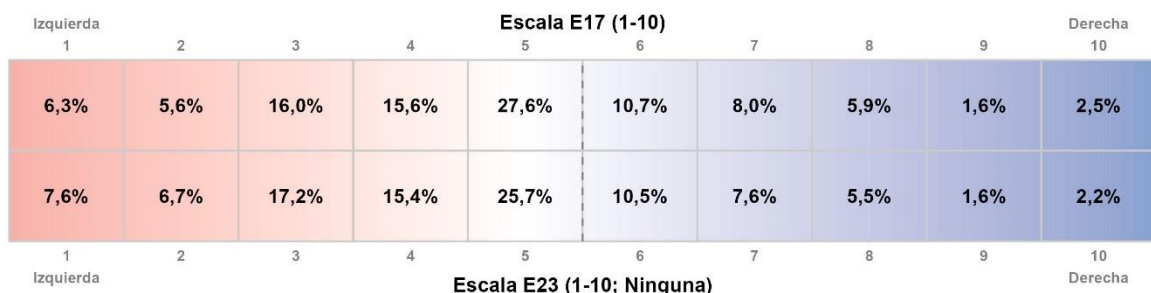


Figura 7. Distribuciones de auto-ubicación ideológica obtenidas con las escalas E17 y E23 de muestras recolectadas en los mismos momentos del tiempo. Detalles en sección S3.1 del material suplementario.

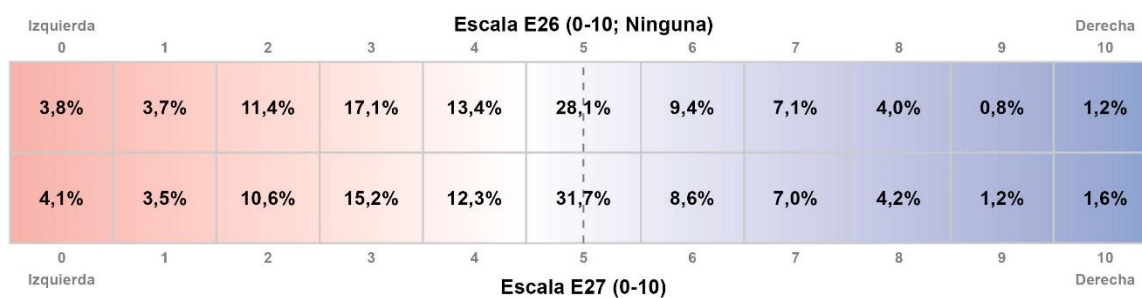


Figura 8. Distribuciones de auto-ubicación ideológica obtenidas con las escalas E26 y E27 de muestras recolectadas en los mismos momentos del tiempo. Detalles en sección S3.1 del material suplementario.

Por último, la Figura 9 compara las escalas E10 y E18, cuyas diferencias de partida son más evidentes pues difieren en el número de puntos que las componen: 7 versus 10. Lo más notable que se observa en la figura es el resultado paradójico de encontrar un mayor porcentaje de encuestados situados en los extremos de las escalas (especialmente en la puntuación 1) en el caso de la escala más extensa. En todo caso, para esta comparación resulta difícil encontrar algún patrón, quizás también consecuencia de que se dispone de menos muestra, ya que ambas escalas sólo coincidieron en un mes: diciembre de 1984.

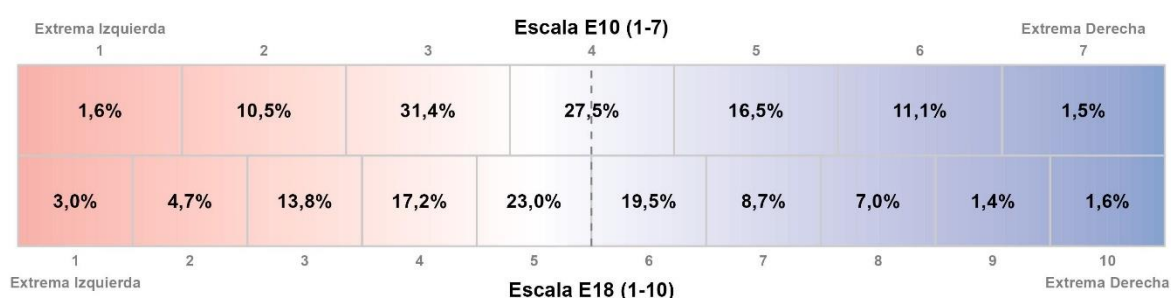


Figura 9. Distribuciones de auto-ubicación ideológica obtenidas con las escalas E10 y E18 de muestras recolectadas en los mismos momentos del tiempo. Detalles en sección S3.1 del material suplementario.

Relaciones de transferencia entre escalas

Los análisis anteriores permiten comparar estimaciones de distribuciones obtenidas en los mismos momentos del tiempo (por lo que teóricamente están aproximando la misma distribución subyacente) con diferentes escalas, pero obtenidas a partir de cuestionarios y muestras diferentes. En este apartado profundizamos en el análisis eliminando el efecto de muestras y cuestionarios diferentes al comparar cómo las mismas personas se sitúan en dos escalas diferentes en los cinco estudios en los que el CIS ha realizado el ejercicio de utilizar en el mismo cuestionario dos escalas de medición (una con solo texto y otra numérica, y casi siempre con un número de categorías de respuesta distintas en cada

escala). Nos referimos a los datos disponibles en los estudios: POL1163, en el que se utilizan las escalas E10 y E11; OTE2731, escalas E9 y E25; POL2743A, escalas E8 y E17; POL2743B escalas E8 y E27; y ELE-3198, escalas E16 y E17. En la sección S4 del material suplementario se ofrece un mayor detalle de estos estudios. Medimos las relaciones entre las escalas a partir de las distribuciones cruzadas de las respuestas de auto-ubicación ideológica de los encuestados.

Las Figuras 10 a 14 muestran las matrices de transferencia entre las escalas utilizadas en cada uno de estos estudios. En tales figuras se ha optado por representar las matrices de transferencia fila estandarizadas, incorporando también la información de las distribuciones cruzadas, lo que permitiría reconstruir con facilidad las matrices de composición (Pavía y Aybar, 2020). Las matrices se han organizado ubicando siempre en el eje vertical la escala con opciones de respuesta solo texto y en el eje horizontal la escala numérica. En las representaciones gráficas hemos recurrido a dos elementos visuales para resaltar el peso que cada celda tiene respecto del total, (a mayor tamaño de fuente mayor peso) y la importancia relativa de cada celda dentro de cada fila (a mayor intensidad de color mayor es el porcentaje correspondiente en términos de la distribución fila estandarizada).

Transferencias entre las escalas E11 y E10 (estudio 1163)

El estudio 1163, realizado en julio de 1978, pertenece al bloque “Política (POL)”. En dicho estudio se preguntó por la auto-ubicación ideológica del encuestado en las preguntas 25 y 33. En la pregunta 25 se utilizó la escala E11 (1-7), ofreciendo 9 alternativas de respuesta de tipo texto (ver Figura 10). En la pregunta 33 se utilizó la escala de medición E10 (1-7), pero especificando solo los textos “Extrema izquierda” y “Extrema derecha”, “No sabe” y “No contesta”. La matriz de transferencia entre ambas escalas se muestra en la Figura 10, donde destaca que la mayoría de las respuestas de auto-ubicación ideológica se contabilizan en la diagonal. Este es el resultado teóricamente esperado, pues a pesar de que en la primera pregunta el orden de las respuestas se planteó con el orden invertido (de derecha a izquierda) la diagonal sintetiza las relaciones texto-números teóricas. La cifra, sin embargo, es más baja de lo esperado a priori: sólo el 53,7% de los que se auto-ubican en alguna de las dos escalas se localizan exactamente en la diagonal.

Otros resultados también destacan. Un 44% (22,7%) de los respondientes que marcaron la opción “No acepta esta distinción” (“NS/NC”) en escala de texto sí que se posicionó en la escala numérica, circunstancia que apunta hacia la conclusión una tasa mayor de auto-posicionamiento cuando las escalas son numéricas. De hecho, en la escala numérica se recolectaron un 11% más de respuestas sustantivas que en la escala de texto. La cifra más alta de coherencia (73,5%) se observa entre los encuestados que optaron por elegir “Centro”, aun así, el 10,6% se auto-ubicó en el 5 (que

correspondería con “Centro derecha”) y el 6% seleccionó el 3 (equivalente a “Centro izquierda”), con un 14% seleccionando otras opciones.

Escala E10 (1-7)

		1	2	3	4	5	6	7	N.S.	N.C.	
Escala E11 (1-7)	Extrema Izquierda	12 (0,545)	6 (0,273)	3 (0,136)		1 (0,046)					22
	Izquierda	5 (0,020)	142 (0,561)	59 (0,233)	10 (0,040)	13 (0,051)	8 (0,032)	1 (0,004)	8 (0,032)	7 (0,028)	253
	Centro Izquierda	1 (0,007)	5 (0,033)	101 (0,664)	23 (0,151)	16 (0,105)	1 (0,007)	1 (0,007)	2 (0,013)	2 (0,013)	152
	Centro		1 (0,005)	13 (0,060)	158 (0,735)	23 (0,107)	3 (0,014)	1 (0,005)	10 (0,047)	6 (0,028)	215
	Centro Derecha		1 (0,018)	4 (0,063)	8 (0,127)	39 (0,619)	7 (0,111)	1 (0,018)	2 (0,032)	1 (0,016)	63
	Derecha	1 (0,017)	4 (0,068)	4 (0,068)	4 (0,068)	8 (0,136)	26 (0,441)		8 (0,136)	4 (0,068)	59
	Extrema Derecha		1 (0,143)				1 (0,143)	5 (0,714)			7
	No acepta esta dist.	5 (0,026)	11 (0,058)	19 (0,100)	36 (0,189)	8 (0,042)	3 (0,016)	1 (0,005)	53 (0,279)	54 (0,284)	190
	NS/NC		5 (0,023)	9 (0,041)	26 (0,118)	8 (0,036)	2 (0,009)		96 (0,436)	74 (0,336)	220
		24	176	212	265	116	51	10	179	148	1181

Figura 10. Estudio POL1163. Comparativa de respuestas utilizando las escalas de medición E11 (1-7), solo texto, y E10 (1-7), escala numérica.

Transferencias entre las escalas E9 y E25 (estudio 2731)

En el estudio 2731 (septiembre de 2017), perteneciente al bloque “Estudios Otros Temas” (EOT) y que tiene por título “Actitudes ante la discriminación por origen racial o étnico”, se preguntó por la auto-ubicación ideológica en dos ocasiones (preguntas 32 y 33). En la pregunta 32 se utilizó la escala E9 (1-6), con 8 alternativas de respuesta de tipo texto (ver Figura 11). En la pregunta 33 se utilizó la escala numérica E25 (0-10), ofreciendo 13 alternativas de respuesta incluyendo “No sabe” y “No contesta”, y con los extremos de la escala etiquetados de la misma manera que los extremos de la escala E9. Al igual que en la Figura 10, en la Figura 11 de nuevo se observa una fuerte coherencia teórica entre las respuestas de los encuestados en ambas clasificaciones. La mayoría de las respuestas están clasificadas en el entorno de la “diagonal”. Destaca, no obstante, que el 24% (20%) de los encuestados que marcaron en la pregunta 32 “No sabe” (“No contesta”) seleccionaron la puntuación 5 en la pregunta 33. Esto pone de manifiesto que ante la ausencia de la opción “Centro”, una parte de los encuestados prefiere no responder a auto-asignarse el adjetivo “izquierda” o “derecha”.

		Escala E25 (0-10)														
		0	1	2	3	4	5	6	7	8	9	10	N.S.	N.C.		
Escala E9 (1-6)	Extrema Izquierda	12 (0,500)	7 (0,282)	5 (0,208)												24
	Izquierda	17 (0,029)	38 (0,064)	190 (0,322)	206 (0,349)	92 (0,156)	36 (0,061)	2 (0,003)	1 (0,002)	1 (0,002)			4 (0,007)	3 (0,005)	590	
	Centro Izquierda	1 (0,002)	1 (0,002)	18 (0,032)	129 (0,230)	271 (0,482)	129 (0,230)	2 (0,004)	1 (0,002)	1 (0,002)			4 (0,007)	5 (0,009)	562	
	Centro Derecha					4 (0,009)	157 (0,360)	203 (0,466)	51 (0,117)	13 (0,030)	2 (0,005)	2 (0,005)	3 (0,007)	1 (0,002)	436	
	Derecha			1 (0,005)		1 (0,005)	14 (0,067)	33 (0,158)	78 (0,373)	64 (0,306)	7 (0,033)	7 (0,033)	4 (0,019)		209	
	Extrema Derecha						1 (0,077)			2 (0,154)		8 (0,615)	1 (0,077)	1 (0,077)	13	
	N.S.	1 (0,002)			4 (0,009)	6 (0,014)	105 (0,241)	4 (0,009)	1 (0,002)			1 (0,002)	283 (0,651)	30 (0,069)	435	
	N.C.			6 (0,012)	5 (0,010)	12 (0,024)	100 (0,196)	10 (0,020)	2 (0,004)	1 (0,002)		1 (0,002)	21 (0,041)	351 (0,690)	509	
		31	46	220	344	386	542	254	134	82	9	19	320	391	2778	

Figura 11. Estudio OTE2731. Comparativa de respuestas utilizando las escalas de medición E9 (1-6), solo texto, y E25 (0-10), escala numérica.

Transferencias entre las escalas E8 y E17 y E8 y E27 (estudios 2743 A y B)

En noviembre de 2007 el CIS realizó un estudio con el objetivo de medir los efectos que generan los formatos de las preguntas. Este examen constó en dos estudios independientes (2743A y 2743B), ya que se aplicó a dos muestras distintas. En ambos se utilizó la escala E8 (1-5), con 7 alternativas de respuesta en una escala de texto (ver Figuras 12 y 13), junto a una escala numérica. En el primero se utilizó la escala E17 (1-10) y en el segundo la escala E27 (0-10).

		Escala E17 (1-10)													
		1	2	3	4	5	6	7	8	9	10	N.S.	N.C.		
Escala E8 (1-5)	Extrema Izquierda	13 (0,812)	1 (0,262)	1 (0,462)	1 (0,002)										16
	Izquierda	40 (0,061)	77 (0,117)	286 (0,436)	182 (0,277)	54 (0,082)	2 (0,003)					9 (0,014)	6 (0,009)	656	
	Centro		3 (0,006)	11 (0,023)	44 (0,094)	259 (0,551)	72 (0,153)	27 (0,057)	13 (0,028)		1 (0,002)	20 (0,043)	20 (0,043)	470	
	Derecha			1 (0,004)	1 (0,004)	14 (0,057)	58 (0,237)	77 (0,314)	55 (0,224)	13 (0,053)	17 (0,069)	3 (0,012)	6 (0,024)	245	
	Extrema Derecha								3 (0,273)	2 (0,182)	5 (0,465)	1 (0,091)		11	
	N.S.	1 (0,003)	3 (0,010)	10 (0,034)	7 (0,024)	39 (0,132)	10 (0,034)	5 (0,017)				181 (0,614)	39 (0,132)	295	
	N.C.	1 (0,003)	1 (0,003)	6 (0,021)	10 (0,035)	52 (0,181)	7 (0,024)	3 (0,010)	1 (0,003)			22 (0,076)	185 (0,642)	288	
	55	85	315	245	418	149	112	72	15	23	236	256	1981		

Figura 12. Estudio POL2743A. Comparativa de respuestas utilizando las escalas de medición E8 (1-5), solo texto, y E17 (1-10), escala numérica.

Como en los dos casos anteriores, la mayoría de encuestados relacionaron (Extrema) Izquierda con los valores numéricos inferiores y (Extrema) Derecha con los valores numéricos superiores, empero no utilizarse el término “Extrema” en la escala numérica. Asimismo, también en estos estudios una parte de los encuestados prefirieron evitar etiquetas. A pesar de estar disponible la opción “Centro” en la escala E8 (1-5), tanto en el estudio 2743A como en el estudio 2743B una parte de los encuestados seleccionaron simultáneamente 5 en la escala numérica y “No sabe” (“No contesta”) en la escala de texto. Un 13,2% y un 18,1%, respectivamente en el primer caso, y un 18,2% y un 19,4% en el segundo. Esto podría indicar que hay un porcentaje de encuestados que, o bien no les gustan las etiquetas o bien identifican el 5 como no posicionamiento ideológico.

Escala E27 (0-10)

		0	1	2	3	4	5	6	7	8	9	10	N.S.	N.C.		
Escala E8 (1-5)	Extrema Izquierda	11 (0,478)	3 (0,130)	2 (0,087)	2 (0,087)	1 (0,043)								4 (0,174)	23	
	Izquierda	25 (0,036)	40 (0,058)	141 (0,204)	237 (0,342)	160 (0,231)	61 (0,088)	2 (0,003)	1 (0,001)		1 (0,001)	1 (0,001)	12 (0,017)	11 (0,016)	692	
	Centro	1 (0,002)		5 (0,011)	26 (0,057)	37 (0,082)	284 (0,627)	42 (0,093)	27 (0,060)	3 (0,007)		3 (0,007)	9 (0,029)	16 (0,035)	453	
	Derecha				4 (0,016)		12 (0,049)	76 (0,311)	68 (0,279)	45 (0,184)	13 (0,053)	20 (0,082)	4 (0,016)	2 (0,008)	244	
	Extrema Derecha					1 (0,167)			1 (0,167)	1 (0,167)	1 (0,167)	1 (0,167)	1 (0,167)		6	
	N.S.	2 (0,007)	1 (0,003)	4 (0,014)	8 (0,027)	9 (0,031)	53 (0,182)	9 (0,031)	4 (0,014)	3 (0,019)	1 (0,003)			169 (0,579)	29 (0,099)	292
	N.C.	5 (0,016)		3 (0,011)	7 (0,028)	10 (0,037)	53 (0,194)	4 (0,015)	5 (0,018)			2 (0,007)	13 (0,048)	171 (0,626)	273	
		44	44	155	284	218	463	133	106	52	16	27	208	233	1983	

Figura 13. Estudio POL2743B. Comparativa de respuestas utilizando las escalas de medición E8 (1-5), solo texto, y E27 (0-10), escala numérica.

Transferencias entre las escalas E10 y E16 (estudio 3198)

El quinto y último estudio en los que el CIS ha preguntado sobre la auto-ubicación ideológica en dos preguntas ha sido el ELE3198, realizado en noviembre de 2017 y titulado “Preelectoral de Cataluña. Elecciones autonómicas de 2017“. En la pregunta 23 se utilizó la escala de medición E10 (1-10), ya comentada anteriormente. En la pregunta 30 se utilizó la escala E16 (1-7), ofreciendo 10 alternativas de respuesta textuales (ver Figura 14). Nuevamente se observa cómo los respondientes (en este caso de la comunidad autónoma de Cataluña) relacionan las opciones de respuesta en ambas preguntas como se esperaría teóricamente: la mayoría de las repuestas se sitúan en la “diagonal”. En concreto, el 64% de respondientes que seleccionaron la opción “Centro” se posicionaron en el número 5 de la escala numérica 1-10, circunstancia que refuerza la hipótesis que afirma que independientemente de que el centro geométrico de la escala sea 5 o 5,5, la gran mayoría de los respondientes identifican el 5 como punto central en escalas con valor máximo 10.

Escala E17 (1-10)

		1	2	3	4	5	6	7	8	9	10	N.S.	N.C.	
Escala E16 (1-7)	Extrema Izquierda	76 (0,916)	5 (0,060)	1 (0,012)		1 (0,012)								83
	Izquierda	364 (0,316)	225 (0,195)	329 (0,286)	109 (0,095)	66 (0,057)	11 (0,010)	8 (0,007)	21 (0,018)	9 (0,007)	2 (0,002)	5 (0,005)	2 (0,002)	1151
	Centro Izquierda	48 (0,073)	43 (0,065)	172 (0,261)	195 (0,295)	147 (0,223)	26 (0,039)	10 (0,015)	16 (0,024)		2 (0,003)		1 (0,002)	660
	Centro	9 (0,017)	3 (0,006)	13 (0,025)	36 (0,068)	338 (0,640)	75 (0,142)	27 (0,051)	16 (0,030)	2 (0,004)	4 (0,008)	2 (0,004)	3 (0,006)	528
	Centro Derecha	2 (0,008)			2 (0,008)	45 (0,176)	58 (0,227)	77 (0,302)	50 (0,196)	8 (0,031)	13 (0,051)			255
	Derecha	2 (0,016)	3 (0,023)			11 (0,086)	9 (0,070)	25 (0,195)	29 (0,227)	10 (0,078)	34 (0,266)	5 (0,039)		128
	Extrema Derecha	2 (0,200)				1 (0,190)	1 (0,100)				5 (0,500)			10
	N.S.	3 (0,050)	2 (0,033)	1 (0,017)	4 (0,087)	16 (0,297)	1 (0,017)	4 (0,067)	3 (0,055)		2 (0,033)	19 (0,317)	5 (0,083)	60
	N.C.	2 (0,044)	1 (0,022)	4 (0,089)		16 (0,356)	2 (0,044)	1 (0,022)	1 (0,022)	2 (0,044)	3 (0,067)	2 (0,044)	11 (0,244)	45
	Ninguna	4 (0,088)	2 (0,025)	3 (0,038)	2 (0,038)	49 (0,613)	1 (0,013)	2 (0,025)	2 (0,025)			9 (0,112)	5 (0,062)	80
		512	284	523	349	690	184	154	138	30	66	43	27	3000

Figura 14. Estudio ELE3198. Comparativa de respuestas utilizando las escalas de medición E16 (1-7), solo texto, y E17 (1-10), escala numérica.

Discusión de resultados

El análisis de las matrices de transferencia de estos cinco estudios, junto a los resultados observados en la sección 5, permite inferir una importante batería de resultados: i) la ausencia de la opción “Centro” aumenta la tasa de no respuesta; ii) una parte relevante de los encuestados, que no se auto-ubican en ninguna opción cuando se les ofrece una escalas de texto, sistemáticamente se auto-ubican en el 5 en las escalas numéricas, actuando este valor como punto de atracción para los que no tienen ideología o prefieren no etiquetarse; iii) en las escalas numéricas 1-10 el Centro se sitúa en el 5, independientemente de que el centro geométrico real sea 5,5; iv) cuando se utilizan escalas 1-7, la mayoría de los encuestados relacionan los pares número-texto como se esperaría teóricamente; v) cuando no se ofrecen las opciones “Centro izquierda” y/o “Centro derecha” los respondientes tienden a posicionarse en opciones más extremas; vi) cuando en el enunciado de la pregunta no se especifica que el 1 se corresponde con “Extrema izquierda” y el 10 con “Extrema derecha”, sino que se indica “Izquierda” y “Derecha”, el encuestado de izquierdas, y no de extrema izquierda, tiende a situarse en el 1.

La lista de resultados anteriores lleva a formularse algunas preguntas adicionales, tales como si ¿existe una mayor tasa de respuesta en las escalas numéricas?, ¿qué tasa de respuesta ofrecen los estudios con solo texto?, y si esta ¿es distinta en los estudios con escalas numéricas? En este punto, el análisis de los datos disponibles nos lleva a situarnos más cerca de la tesis de McRoy y Pancratz (2020) que de la de Krosnick (1991) al observar mayores tasas de respuesta cuando se emplean escalas numéricas. Analizando los 1454 estudios en que se dispone de microdatos se observa que en los 12

donde se utilizaron escalas solo texto la tasa de respuesta agregada para la pregunta auto-ubicación ideológica fue del 69,2% (N = 21.425), cuando, por el contrario, en los 1442 estudios en los que se ha utilizado escala numérica la tasa de respuesta agregada es del 75,8% (N=3.587.577).

Esta comparación invita a profundizar en las tasas de respuesta, estudiando si existen o no diferencias entre tasas de respuesta dependiendo del tipo de estudio o del tipo de escala utilizada, entendiendo por no respuesta que el encuestado escoja una opción de respuesta fuera de las opciones de la propia escala Likert. Por un lado, las tasas de respuesta por tipo de estudio son: i) BAR 77,0% (N=1.069.687 en 397 estudios); ii) ELE 79,9% (N=920.078 en 374 estudios); iii) ICC 81,4% (N=229.723 en 128 estudios); iv) OTE 70,2% (N=1.047.000 en 390 estudios); y POL 73,6% (N=342.514 en 165 estudios). Por otro lado, las tasas de respuesta según el tipo de escala utilizada son: i) 1-7 69,0% (N=200.440 en 41 estudios); ii) 1-10 77,0% (N=3.256.734 en 1367 estudios); y iii) 0-10 76,1% (N=145.086 en 43 estudios). El resto de tipos de escala aglutinan una cantidad de estudios demasiado pequeña para que sus tasas sean comparables.

Atendiendo a la tipología de estudios, destacan las cifras de ICC, con 81,4%, y OTE, con 70,2%. El primero siempre emplea encuestas telefónicas, mientras el segundo suele recoger información habitualmente no percibida como de temas políticos. Estas circunstancias podrían estar detrás de estos resultados. En todo caso, sería preciso un análisis más pormenorizado de esta cuestión. Por otro lado, atendiendo a las tasas de respuesta según tipo de escala, los resultados aparentemente apuntan a que las posibilidades de respuesta se incrementan con un mayor número de opciones. Los diferentes momentos temporales en que se utilizaron mayoritariamente ambos tipos de escala, sin embargo, no permiten descartar que la explicación sea otra.

Resumen y conclusiones

En el contexto de las encuestas del CIS y de la medición de la ideología a través de la auto-ubicación ideológica de los encuestados, en esta investigación se han identificado y analizado hasta 28 escalas de medición distintas, agrupadas en 8 atendiendo al número de alternativas de respuesta de que dispone la escala propiamente dicha. Las 28 escalas se han construido después de diferenciar si las opciones de respuesta eran numéricas o solo incluían texto, y si disponían de otras opciones de respuesta más allá del “No sabe” o “No contesta”. El análisis se ha basado en el estudio de casi 2000 estudios que desde 1976 han venido indagando sobre la auto-ubicación ideológica de los españoles y ha abarcado la búsqueda de respuesta a una gran cantidad de hipótesis (preguntas), sobre las cuales será preciso profundizar en el futuro en más detalle y de manera focalizada.

Se ha observado que la escala más utilizada por el CIS ha sido la E17 (1-10), para la cual, en línea con lo apuntado por del Castillo (2004), se ha constatado que el 5 actúa como centro efectivo de la escala. Una gran mayoría de los respondientes identifican el 5 como centro de esta escala a pesar de ser una escala sin un valor central definido. Esta circunstancia produce cierto sesgo de subestimación. También se ha observado que el etiquetado de todas las opciones de respuestas puede desincentivar el posicionamiento de una parte de los encuestados. En general, no obstante, como norma, las relaciones entre las etiquetas políticas y las puntuaciones numéricas coinciden con las teóricamente esperadas.

En términos de tasa de respuesta, no se observan diferencias significativas entre los estudios que utilizan escalas 1-10 y los que utilizan escalas 0-10. Los estudios que utilizan escalas 1-7, sin embargo, arrojan unas tasas de respuesta notablemente inferiores. Por tipo de estudio, observamos que los estudios ICC obtienen las tasas de respuesta más elevadas. Esto puede ser debido al método de realizar la entrevista, ya que estos estudios utilizan entrevistas telefónicas. Sería interesante, por tanto, analizar, en posteriores investigaciones, cómo influye en la tasa de respuesta el modo de administración del cuestionario (por vía telefónica, presenciales o auto-administradas por internet) y la forma en que se presentan las preguntas.

A la vista de todo lo anterior, y a modo de conclusión, y siguiendo la línea propuesta por McRoy y Pancratz (2020), los análisis apuntan a que lo más apropiado para medir la auto-ubicación ideológica de los españoles sería utilizar una escala 0-10, por ser una escala numérica con un número de opciones impar y amplio. Impar para representar el centro correctamente (escala simétrica), y amplio para que el respondiente pueda aproximar su ideología política a un valor numérico distinguiendo cierto grado de centrismo/extremismo. Se trataría de una escala compuesta por 11 valores en los que implícitamente se dispondría de 3 opciones para el centro, 3 opciones para la izquierda y 3 opciones para la derecha, además de los valores 0 y 10 que se deberían acompañar de “Extrema izquierda” y “Extrema derecha”, respectivamente, evitando posibles confusiones detectadas en esta investigación. Estas 11 alternativas de respuesta deberían ser completadas con, además de las opciones tradicionales “No sabe” y “No contesta”, la opción “Apolítico/a” o alguna similar.

Referencias

Alabi, A.T. y M.O. Jelili. 2022. “Clarifying Likert scale misconceptions for improved application in urban studies”, *Quality and Quantity*. <https://doi.org/10.1007/s11135-022-01415-8>

- Aybek, E.C. y C. Toraman. 2022. "How many response categories are sufficient for Likert type scales? An empirical study based on the Item Response Theory", *International Journal of Assessment Tools in Education* 9(2): 534-547.
- Bauer, P.C., P. Barberá, K. Ackermann, y A. Venetz. 2016. "Is the left-right a valid measure of ideology", *Political Behavior* 39: 553–583. <https://doi.org/10.1007/s11109-016-9368-2>
- Bisquerra, R. y N. Pérez-Escoda. 2015. "¿Pueden las escalas Likert aumentar en sensibilidad?", *Revista d'Innovació i Recerca en Educació (REIRE)* 8(2). <https://doi.org/10.1344/reire2015.8.2828>
- Camino, E. 2022. "La no respuesta a la autoubicación ideológica: una distribución social del silencio", *Revista de Investigaciones Políticas y Sociológicas* 21(1): 1-18. <https://doi.org/10.15304/riips.21.1.7695>
- Cano Espiñeira, A. E. 2015. "Dimensionalización de la ideología en la Encuesta Social Europea: efecto de la no respuesta y testado". Universidad de Alicante. Departamento de Sociología II. URI: <http://hdl.handle.net/10045/48739>
- Canto de Gante, A. G., W. E. Sosa, J. B. Ortega, J. Escobar y A. Santillán. 2020. "Escala de Likert: Una alternativa para elaborar e interpretar un instrumento de percepción social". *Revista de la Alta Tecnología y Sociedad* 12(1): 38-45.
- Carroll, R. y H. Kubo. 2021. "Measuring and explaining the complexity of left-right perceptions of political parties", *Electoral Studies* 71. <https://doi.org/10.1016/j.electstud.2021.102310>
- Cummins, R.A. y E. Gullone. 2000. "Why we should not use 5-point Likert scales: The case for subjective quality of life measurement", *Second International Conference on Quality of Life in Cities*, 74-93. Singapore: National University of Singapore.
- Dassonneville, R. 2020. "Replication data for: Change and continuity in the ideological gender gap. A longitudinal analysis of left-right self-placement in OECD countries", *European Journal of Political Research*. <https://doi.org/10.7910/DVN/B5VGCD>
- Del Castillo, P. 2004. "Nota metodológica sobre los indicadores del barómetro del CIS", *Revista Española de Investigaciones Sociológicas (REIS)* 108: 151-178.
- DeVellis, R.F. 2003. "Scale development, theory and applications", *SAGE Publications*.
- Evans, J. 2019. "On scale from 1 to 10, how much do the numbers used in survey scales really matter?", <https://medium.com/pew-research-center-decoded/how-much-do-the-numbers-used-in-survey-scales-really-matter-227d84ab2a13>

- Gestefeld, M., J. Lorenz, N.T. Henschel y K. Boehnke 2022. “Decomposing attitude distributions to characterize attitude polarization in Europe”, *SN Social Sciences* 2: 110. <https://doi.org/10.1007/s43545-022-00342-7>
- Harpe, S.E. 2015. “How to analyze Likert and other rating scale data”. *Currents in Pharmacy Teaching and Learning* 7: 836-850. <http://dx.doi.org/10.1016/j.cptl.2015.08.001>
- Huiping, W. y L. Shing-On. 2017. “Can Likert scales be treated as interval scales? A simulation study”, *Journal of Social Service Research* 43(4): 527-532. <http://dx.doi.org/10.1080/01488376.2017.1329775>
- Jebb, A.T., V. Ng y L. Tay. 2021. “A Review of key Likert scale. Development advances: 1995-2019”, *Frontiers in Psychology* 12: 637547. <http://dx.doi.org/10.3389/fpsyg.2021.637547>
- Joshi, A., S. Kale, S. Chandel y D.K. Pal. 2015. “Likert scale: explored and explained”, *British Journal of Applied Science and Technology* 7(4): 396-403. <https://doi.org/10.9734/BJAST/2015/14975>
- Kroh, M. 2007. “Measuring left-right political orientation: The choice of response format”, *Public Opinion Quarterly*, 71(2): 204-220. <https://doi.org/10.1093/poq/nfm009>
- Krosnick, J.A. 1991. “Response strategies for coping with the cognitive demands of attitude measures in surveys”, *Applied Cognitive Psychology* 5: 213-236.
- Lelkes, Y. 2016. “Mass polarization: manifestations and measurements”, *Public Opinion Quarterly* 80(S1): 392–410, <https://doi.org/10.1093/poq/nfw005>
- Likert, R. 1932. “A technique for the measurement of attitudes”, *Archives of Psychology* 140: 1-55.
- Matas, A. 2018. “Diseño del formato de escalas tipo Likert: un estado de la cuestión”, *Revista Electrónica de Investigación Educativa* 20(1): 38-47. <https://doi.org/10.24320/redie.2018.20.1.1347>
- McRoy, M. y S. Pancratz. 2020. “Small changes in survey scales can matter when measuring political ideology in Europe”, <https://medium.com/pew-research-center-decoded/small-changes-in-survey-scales-can-matter-when-measuring-political-ideology-in-europe-4a10d9a015c5>
- Mircioiu, C. y J. Atkinson. 2017. “A comparison of parametric and non-parametric methods applied to a Likert scale”, *Pharmacy* 5(2): 26. <http://dx.doi.org/10.3390/pharmacy5020026>

- Nilsson, A. y A. Erlandsson. 2015. "The moral foundations taxonomy: structural validity and relation to political ideology in Sweden", *Personality and Individual Differences* 76: 28-32. <http://dx.doi.org/10.1016/j.paid.2014.11.049>
- Pavía, J.M. y Aybar, C. 2018. "Field rules and bias in random surveys with quota samples. An assessment of CIS surveys", *SORT* 42(2): 183-206. <http://dx.doi.org/10.2436/20.8080.02.74>
- Pavía, J.M. y Aybar, C. 2020. "La movilidad electoral en las elecciones de 2019 en la Comunitat Valenciana", *Debats. Revista de Cultura, Poder y Sociedad* 134(1): 27-51. <https://doi.org/10.28939/iam.debats.134-1.3>
- Sánchez Medero, G. y R. Sánchez Medero. 2009. "Las orientaciones ideológicas de los ciudadanos en Europa", *Papel Político Bogotá (Colombia)* 14(2): 645-667.
- Sánchez, P.A. 2017. "Construcción de una escala ideológica política en el contexto costarricense a partir de un trabajo realizado en Córdoba, Argentina". *Acta de Investigación Psicológica* 7: 2747-2754. <http://dx.doi.org/10.1016/j.aippr.2017.06.005>
- Schwartz, B. 2005. *Por qué más es menos. La tiranía de la abundancia*. Taurus.
- Sinclair, V. M. 2022. "Development of a new measure of political ideology", *Electronic Thesis and Dissertation Repository* 8449. <https://ir.lib.uwo.ca/etd/8449>
- Stefkovich, Á. 2022. "Are scale direction effects the same in different survey modes? Comparison of a face-to-face, a telephone, and an online survey experiment", *Field Methods* 34(3): 206-222. <https://10.1177/1525822X221105940>
- Taherdoost, H. 2019. "What is the best response scale for survey and questionnaire; Review of different lengths of rating scale / attitude scale / Likert scale". *International Journal of Academic Research in Management* 8(1): 1-10.
- Todosijevic, B. 2016. "Left-right ideology: its meaning and effects on party preferences in Serbia", *Социолошки преглед* 50(2): 161-178.
- Weijters, B., K. Millet y E. Cabooter. 2021. "Extremity in horizontal and vertical Likert scale format responses. Some evidence on how visual distance between response categories influences extreme responding", *International Journal of Research in Marketing* 38: 85-103.
- Willits, F.K., G.L. Theodori y A. E. Luloff. 2016. "Another look at Likert scales", *Journal of Rural Social Sciences* 31(3): 6. <https://egrove.olemiss.edu/jrss/vol31/iss3/6>

Wu, H., y S. O. Leung. 2017. "Can Likert scales be treated as interval scales? A simulation study", *Journal of Social Service Research* 43(4): 527-532. <https://doi.org/10.1080/01488376.2017.1329775>

Zuell, C. y E. Scholz. 2019. "Construct equivalence of left-right scale placement in a cross-national perspective", *International Journal of Sociology* 49(1): 77-95. <https://doi.org/10.1080/00207659.2018.1560982>

Anexo IV

Dataset of the COVID-19 lockdown survey conducted by GIPEyOP in Spain

Pérez, V., Aybar, C. & Pavía, J. M. (2022). *Data in Brief*, 40, 107700.

<https://doi.org/10.1016/j.dib.2021.107700>



Data Article

Dataset of the COVID-19 lockdown survey conducted by GIPEyOP in Spain

Virgilio Pérez^a, Cristina Aybar^a, Jose M. Pavía^{b,*}^a GIPEyOP, Department of Applied Economics, University of Valencia, Spain^b GIPEyOP, UMMICS, Department of Applied Economics, University of Valencia, Spain

ARTICLE INFO

Article history:

Received 8 October 2021

Revised 8 November 2021

Accepted 8 December 2021

Available online 10 December 2021

Keywords:

SARS-CoV-2

Confinement

Online surveys

Chain sampling

Social impacts

Political assessment



ABSTRACT

This article introduces a dataset that captures relevant information about the living conditions, feelings, and habits of residents in Spain during ninety nine days of home confinement. This and other measures, imposed by the Government of Spain to mitigate the impact of the pandemic on the population, have brought with them important economic, labor, and social changes, which have been accompanied by various modifications (some only temporary) in Spaniards habits and behaviours.

Data collection was carried out through the implementation of a questionnaire with 33 questions, which was sent by email to the collaborators of GIPEyOP (Elections and Public Opinion Research Group from the University of Valencia). These collaborators, in turn, forwarded the questionnaire to their acquaintances using email and social networks, mainly WhatsApp, Facebook, and Twitter. This non-probabilistic methodology has generated a total of 8387 valid responses.

The resulting dataset may be (re)used by sociologists, political scientists, economists, or psychologists, among others, to identify how household chores were distributed among family members during the lockdown, what impact the confinement had on the labor performance of workers, the extent of teleworking and on some (physical and psychological) health issues linked to the confinement, including relationships with the place of residence during confinement.

* Corresponding author.

E-mail address: pavia@uv.es (J.M. Pavía).Social media:  (V. Pérez),  (C. Aybar)

The data also provides information on how social networks spread geographically or what Spaniards thought of the management of the crisis by local, national, and international authorities.

© 2021 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Social Science, Sociology, Political Science, Health, Economy
Specific subject area	Social Science (general), Public Opinion, Political Science, Health, Economics
Type of data	Table (spreadsheet)
How the data were acquired	Data were collected using a self-administered online questionnaire. The questionnaire is provided in Spanish (original) and English (translated) as a supplementary file (word format). A snowball or chain sampling method was used to recruit respondents.
Data format	Raw
Description of data collection	The survey was carried out during the period of home confinement decreed by the Spanish government from mid-March, 2020, motivated by the evolution of the COVID-19 pandemic. The survey data were collected over seventeen days (between 28th April and 14th May, 2020).
Data source location	Country: Spain
Data accessibility	Data file (spreadsheet) is supplied as supplementary material with this article.

Value of the Data

- This dataset offers information on various aspects, not directly observable, related to the Spanish population, how they lived and felt as well as their perceptions while in lockdown during the COVID-19 pandemic.
- Social scientists, including sociologists, economists, and political scientists could use this data to assess gender theories of behavior within the home in a scenario of the blurring of traditional gender roles.
- Economists and psychologists could use these data to assess issues related to the performance of workers (and students) under stressful conditions and the extent of teleworking.
- These data are also of value for political scientists to assess governments and the evolution of public opinion from a comparative perspective.

1. Data Description

This document describes the data collected through a survey conducted between 28/04/2020 and 14/05/2020 on the Spanish population. The dataset contains 122 variables and a total of 8387 rows. The questionnaire, available as supplementary material to this article in Spanish and translated to English, is divided into 6 sections and consists of 33 questions.

Table 1 shows a brief description of the 122 variables available in the dataset. The mismatch between the number of questions (33) and of variables (122) comes from the fact that there are many questions for which more than a variable is extracted. For instance, in one of the questions of the second section of the questionnaire, the respondent is asked to list the outdoor spaces its living place during confinement has (variable OUTDOOR). The respondent could choose up to five available options. This generates 5 variables from a single question, a circumstance that is repeated several times throughout the questionnaire.

Table 1 (continued)

Section	Question code	Question	Values
III	3001	Employment situation of the respondent	<ol style="list-style-type: none"> 1. I am salaried and telework. I work from home 2. I am a salaried employee and I leave home to work 3. I am self-employed and telework. I work from home 4. I am self-employed and I leave home to work 5. I am temporarily laid-off from work 6. I am a salaried employee with no possibility of working 7. I am self-employed without the possibility of working 8. I have been fired during the state of alarm period 9. Sick leave/pregnancy 10. I am unemployed or on leave of absence 11. Retired 12. Student 13. Unpaid work at home 14. Other
	3001A	Do you feel that your productivity at work has been affected by the new situation? <i>Note: only to be answered by respondents who chose 1, 2, 3 or 4 in question 3001.</i>	<ol style="list-style-type: none"> 1. Yes 2. No
	3001B	What is your experience of working at home? <i>Note 1: only to be answered by respondents who chose 1 or 3 in question 3001.</i> <i>Note 2: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. I am making better use of my time than in my workplace 2. It is difficult to reconcile work and family life 3. I would not mind continuing to telework 4. I prefer to commute to the workplace 5. I would like to alternate between the two options
	3001C	Do you think your work will be affected when the state of alarm ends? <i>Note 1: only to be answered by respondents who chose 1 or 3 in question 3001.</i> <i>Note 2: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. Yes, because of a lack of economic activity due to the crisis 2. Yes, due to staff cuts 3. Yes, due to salary cuts 4. Yes, because of having to help in the family environment and having underperformed at work 5. No, everything will stay the same
	3001D	Do you think your work will be affected when the state of alarm ends? <i>Note 1: only to be answered by respondents who chose 2 or 4 in question 3001.</i> <i>Note 2: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. Yes, because of a lack of economic activity due to the crisis 2. Yes, due to staff cuts 3. Yes, due to salary cuts 4. No, everything will stay the same
	3001E	Do you feel that your productivity in your studies has been affected by the new situation? <i>Note: only to be answered by respondents who answered 12 in question 3001.</i>	<ol style="list-style-type: none"> 1. Yes 2. No

(continued on next page)

Table 1 (continued)

Section	Question code	Question	Values
	3001F	If you have experienced difficulties in continuing your studies during this confinement, could you indicate which ones? <i>Note 1: only to be answered by respondents who chose 12 in question 3001.</i> <i>Note 2: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. I do not have a computer now 2. I have to share a computer 3. Internet connection problems 4. I do not have a quiet space at home to study 5. Lack of motivation 6. Lack of time to combine studies with obligations at home/family 7. Disorganisation of online classes 8. Lack of material 9. None of the above
	3001G	Which of the following statements do you identify with? <i>Note 1: only allowed to be answered by respondents who answered 12 in question 3001.</i> <i>Note 2: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. This situation motivates me to push forward with my course 2. I have lost my enthusiasm for the course 3. I can prepare the subjects with the resources provided by the teacher 4. I am organised and I will succeed 5. I am not used to studying on my own and I will not succeed
IV	4001	Weekly frequency of the corresponding task (15 tasks and 2 moments in time). <i>Note: tasks and moments in time are shown in Fig. 1.</i>	0 times to 7 times; No proceed to response
V	5001	Fear of leaving the home	<ol style="list-style-type: none"> 1. I have not been out for the entire period of confinement, and I am afraid to do so 2. I have not been out for the whole period of confinement, but I am not afraid to do so 3. I go out to do chores (walking the dog, shopping, work, care...) and I do it with fear 4. I have gone out just enough to shop and/or work and I am not afraid 5. I have gone out whenever I can, and I have no fear
	5002	Weekly frequency of exercise at home	<ol style="list-style-type: none"> 1. 7 days 2. Between 4 and 6 days 3. Between 2 and 3 days 4. Once 5. Never
	5003	Have you been monitoring your diet in any way?	<ol style="list-style-type: none"> 1. Yes, I am eating less 2. No, being at home I eat more often 3. I eat the same as usual
	5004	Have you had or do you have symptoms related to coronavirus?	<ol style="list-style-type: none"> 1. Yes, I tested positive with symptoms and isolated myself at home 2. Yes, I tested positive while asymptomatic and isolated myself at home 3. Yes, I have spent days in hospital 4. Yes, I have been in a hotel on medication 5. I think so, in these days of confinement

(continued on next page)

Table 1 (continued)

Section	Question code	Question	Values
			6. I think so, before I was confined 7. I have/had no symptoms, but I have not been tested
	5005	In general, are you sleeping as well as before the current crisis?	1. Yes 2. No
VI	6001	Valuation of the national government in terms of the health crisis management.	0 (very bad) to 10 (very good)
		Valuation of the national government in terms of the economic crisis management.	0 (very bad) to 10 (very good)
		Valuation of the regional government in terms of the health crisis management.	0 (very bad) to 10 (very good)
		Valuation of the regional government in terms of the economic crisis management.	0 (very bad) to 10 (very good)
		Valuation of the local government in terms of the health crisis management.	0 (very bad) to 10 (very good)
		Valuation of the local government in terms of the economic crisis management.	0 (very bad) to 10 (very good)
	6002	PSOE and UP act in the same direction.	1 (totally disagree) to 5 (total agreement)
		The government is reporting transparently.	1 (totally disagree) to 5 (total agreement)
		The government is being resolute.	1 (totally disagree) to 5 (total agreement)
		The government declared a state of alarm at the right time.	1 (totally disagree) to 5 (total agreement)
		I have confidence in the government's management of the health crisis.	1 (totally disagree) to 5 (total agreement)
		I have confidence in the government's handling of the economic crisis.	1 (totally disagree) to 5 (total agreement)
		The government is getting international recognition for its handling of this pandemic.	1 (totally disagree) to 5 (total agreement)
	6003	How would you rate PP performance in this crisis?	0 (very bad) to 10 (very good)
		How would you rate VOX performance in this crisis?	0 (very bad) to 10 (very good)
	6004	If PP were in government, what do you think their (health) management of this crisis would have been like?	1. Better 2. Same 3. Worse
		If PP were in government, what do you think their (economic) management of this crisis would have been like?	1. Better 2. Same 3. Worse
	6005	If VOX were in government, what do you think their (health) management of this crisis would have been like?	1. Better 2. Same 3. Worse
		If VOX were in government, what do you think their (economic) management of this crisis would have been like?	1. Better 2. Same 3. Worse
	6006	How would you assess the overall response to the COVID-19 crisis in EU?	0 (very bad) to 10 (very good)
		How would you assess the overall response to the COVID-19 crisis in Germany?	0 (very bad) to 10 (very good)
		How would you assess the overall response to the COVID-19 crisis in China?	0 (very bad) to 10 (very good)

(continued on next page)

Table 1 (continued)

Section	Question code	Question	Values
		How would you assess the overall response to the COVID-19 crisis in the United States?	0 (very bad) to 10 (very good)
		How would you assess the overall response to the COVID-19 crisis in France?	0 (very bad) to 10 (very good)
		How would you assess the overall response to the COVID-19 crisis in the Netherlands?	0 (very bad) to 10 (very good)
		How would you assess the overall response to the COVID-19 crisis in Italy?	0 (very bad) to 10 (very good)
		How would you assess the overall response to the COVID-19 crisis in Portugal?	0 (very bad) to 10 (very good)
		How would you assess the overall response to the COVID-19 crisis in the United Kingdom?	0 (very bad) to 10 (very good)
		How would you assess the overall response to the COVID-19 crisis in Sweden?	0 (very bad) to 10 (very good)
	6007	Could you tell me which party you voted for in the last General Election?	1. List of parties in Table A2 (Appendix file) 2. Others 3. I was not old enough to vote 4. Abstention 5. I voted blank
	6008	If a congressional election were held today, which party would you vote for?	1. List of parties in Table A2 (Appendix file) 2. I would not vote
		Time taken to complete the questionnaire	Numbers of seconds taken
		Time needed to complete section I	Numbers of seconds taken
		Time needed to complete section II	Numbers of seconds taken
		Time needed to complete section III	Numbers of seconds taken
		Time needed to complete section IV	Numbers of seconds taken
		Time needed to complete section V	Numbers of seconds taken
		Time needed to complete section VI	Numbers of seconds taken

The dataset and the dictionary of variables are supplied as supplementary material. In the dataset (spreadsheet) two types of missing values can be distinguished: blank cells, corresponding to non responses, and cells with the value N/A (Not Applicable) which refer to those questions not applicable for those surveyed for whom a certain question did not need answering due to their answers in previous questions. For instance, some of the questions in Section III (dedicated to the work/educational environment) depend on which answer is given to the first question of this section (see [Table 1](#)).

In the first section, Section I, which consists of five questions, data are collected on the sociodemographic characteristics of the respondents. For reasons of space, the detail of the values for the PROV variable, the Spanish province to which the respondent's municipality of residence belongs, is provided in Table A1 (Appendix file) and not in [Table 1](#). Section II of the survey, with 6 questions, asks about the conditions of the residence in which the respondent was confined during the lockdown. Section III investigates the employment situation of the respondents. Some of the questions in the third section, made up of 8 questions about the work environment (or studies), depend on which answer is given to the first question in that block (see [Table 1](#)).

Section IV consists of a single question, which constitutes one of the central questions of the questionnaire and has led to the research reported in [1]. [Fig. 1](#) provides a summary of the responses to this question based on mean values. The question asks how often different domestic tasks, 15 examples in total, are carried out weekly at two points in time (before and during confinement), thus generating 30 variables. [Table 2](#) shows the frequency distributions

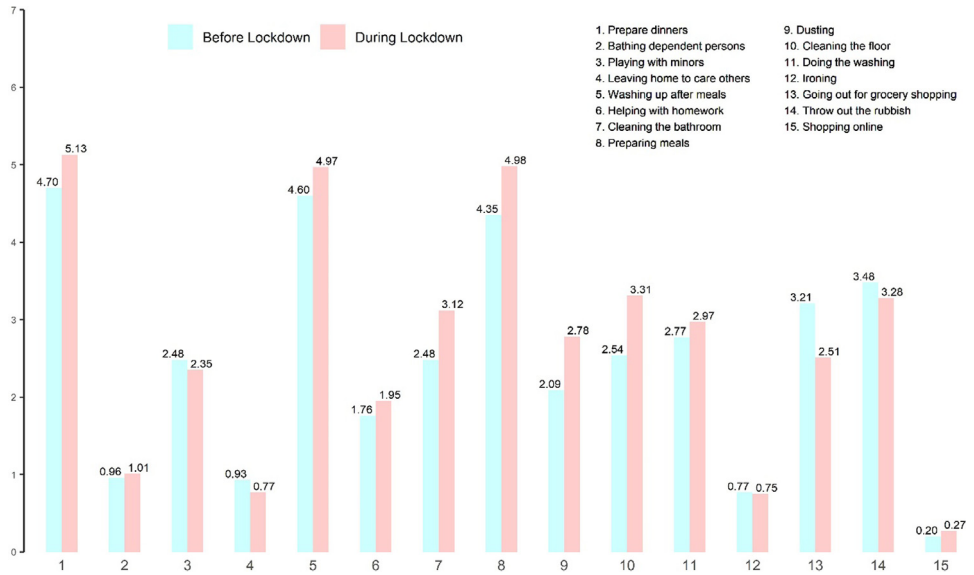


Fig. 1. Weekly frequency (average number of days) of the respondent performing certain household tasks (before and during confinement).

Table 2
Section V questions distribution.

Variable	Category	f (%)
FEAR	I have not been out for the entire period of confinement, and I am afraid to do so.	535 (7.17)
	I have not been out for the whole period of confinement, but I am not afraid to do so.	649 (8.69)
	I go out to do chores (walking the dog, shopping, work, care...) but I am afraid.	2286 (30.62)
	I have gone out just enough to shop and/or work and I am not afraid.	279 (3.74)
	I have gone out whenever I can, but I am afraid.	3717 (49.79)
EXERCISE	7 days	1245 (16.54)
	Between 4 and 6 days	1800 (23.92)
	Between 2 and 3 days	1776 (23.60)
	Once	1144 (15.20)
	Never	1561 (20.74)
DIET	Yes, I am eating less.	1528 (20.30)
	No, being at home I eat more often.	1784 (23.70)
	I eat the same as usual.	4214 (55.99)
COVID	Yes, I tested positive with symptoms and isolated myself at home.	55 (0.82)
	Yes, I tested positive while asymptomatic and isolated myself at home.	8 (0.12)
	Yes, I have spent days in hospital.	22 (0.33)
	Yes, I have been in a hotel on medication.	0 (0.00)
	I think so, during the confinement period.	341 (5.09)
	I think so, before the confinement period.	567 (8.47)
SLEEP	I have/had no symptoms, but I have not been tested.	5705 (85.17)
	Yes	3827 (50.64)
	No	3730 (49.36)

Table 3

Average assessment rating of governments by respondents regarding the management of the health and economic dimensions of the crisis.

Region of residence	Health assessment			Economic assessment			Country	Rating
	National	Regional	Local	National	Regional	Local		
Spain (whole sample)	5.07	5.50	5.65	4.94	5.08	5.19	European Union	3.97
Andalucía	4.83	4.43	5.21	4.71	3.68	4.75	Germany	6.18
Aragón	4.59	4.99	5.38	4.52	4.77	4.97	China	4.97
Canarias	6.68	7.20	6.16	6.85	6.68	5.77	United States	1.60
Cantabria	3.91	4.55	5.86	3.40	4.00	5.20	France	4.61
Castilla-La Mancha	4.25	2.91	4.55	4.03	3.03	4.20	Netherlands	4.45
Castilla y León	5.47	4.38	4.64	5.64	3.86	4.53	Italy	4.12
Cataluña	4.14	4.34	4.69	3.79	3.56	3.92	Portugal	7.21
Ceuta y Melilla	4.38	2.71	2.86	4.50	2.86	2.86	United Kingdom	2.75
Comunidad de Madrid	4.78	4.09	5.64	4.71	3.72	4.93	Sweden	5.66
C. Foral de Navarra	3.59	4.94	3.86	3.63	4.33	3.45		
Comunitat Valenciana	5.18	5.91	5.82	5.06	5.53	5.40		
Extremadura	2.78	3.81	5.48	3.25	3.88	4.78		
Galicia	5.72	4.53	5.19	5.12	3.45	4.23		
Illes Balears	5.25	6.45	6.26	5.32	5.98	5.45		
La Rioja	4.89	5.85	5.06	4.00	4.38	3.73		
País Vasco	5.76	5.83	5.30	5.85	5.47	5.41		
Principado de Asturias	5.29	6.06	5.23	5.22	5.00	4.49		
Región de Murcia	4.89	5.65	5.86	4.44	4.39	5.07		

Table 4

Sample size by autonomous community.

Region	Sample size	Region	Sample size
España	8387	Comunidad de Madrid	606
Andalucía	266	C. Foral de Navarra	21
Aragón	235	Comunitat Valenciana	6014
Canarias	55	Extremadura	27
Cantabria	11	Galicia	100
Castilla-La Mancha	216	Illes Balears	54
Castilla y León	184	La Rioja	26
Cataluña	372	País Vasco	59
Ciudad de Ceuta	2	Principado de Asturias	69
Ciudad de Melilla	7	Región de Murcia	63

associated with the five questions that make up Section V, focused on respondent feelings and daily habits during the lockdown.

The last section, Section VI, is dedicated to the assessment of political management of the pandemic and questions related to the electoral debate. [Table 3](#) summarises the results achieved for some of the variables of this block. Specifically, it offers the assessment, on average, that respondents make about how effectively the government has managed the situation at local, regional, and national levels, and also offers information on the perception that Spaniards have of how the crisis was being managed in other countries.

To complement part of the information contained in the responses collected, [Table 4](#) shows the distribution of the sample size by autonomous communities and [Table 5](#) presents a broad summary of the profile of the respondents in the survey. This explains the composition of the sample in terms of the main socio-economic-demographic characteristics, variables that in conjunction with R.VOTE (see [Table A2](#) in Appendix file) can be used to correct biases by applying calibration or post-stratification techniques.

Table 5Respondent characteristics ($n = 8387$).

Characteristics	Category	Frecuency (%)
Gender	Male	3834 (45.70)
	Female	4553 (54.30)
Age (years)	< 20	66 (0.79)
	20–25	446 (5.32)
	26–30	405 (4.83)
	31–35	531 (6.33)
	36–40	602 (7.18)
	41–45	834 (9.94)
	46–50	908 (10.83)
	51–55	1136 (13.54)
	56–60	1120 (13.35)
	61–65	1065 (12.70)
	66–70	753 (8.98)
	>70	514 (6.13)
	<i>in blank</i>	7 (0.08)
Employment situation	I am salaried and telework. I work from home.	2183 (26.03)
	I am a salaried employee and I leave home to work.	1136 (13.54)
	I am self-employed and telework. I work from home.	301 (3.59)
	I am self-employed and I leave home to work.	248 (2.96)
	I am temporarily laid-off from work.	556 (6.63)
	I am a salaried employee with no possibility of working.	160 (1.91)
	I am self-employed without the possibility of working.	320 (3.82)
	I have been fired during the state of alarm period.	124 (1.48)
	Sick leave/pregnancy.	214 (2.55)
	I am unemployed or on leave of absence.	475 (5.66)
	Retired	1623 (19.35)
	Student	428 (5.10)
	Unpaid work at home	178 (2.12)
Education	Other	338 (4.03)
	<i>in blank</i>	103 (1.23)
	Without studies	13 (0.16)
	Primary education	351 (4.19)
	Secondary education	366 (4.36)
	Job training	984 (11.73)
	Baccalaureate	1023 (12.20)
	University studies	4934 (58.83)
	Doctorate	631 (7.52)
	<i>in blank</i>	85 (1.01)
Residence municipality size (inhabitants)	Less than 2,000 inhabitants	386 (4.60)
	Between 2,001 and 10,000	1018 (12.14)
	Between 10,001 and 50,000	2215 (26.41)
	Between 50,001 and 100,000	794 (9.47)
	Between 100,001 and 400,000	839 (10.00)
	Between 400,001 and 1,000,000	1697 (20.23)
	More than 1,000,000 de inhabitants	1255 (14.96)
	<i>in blank</i>	183 (2.18)
Home size (squared meters)	Less than de 35 m ²	44 (0.52)
	Between 35 and 50 m ²	362 (4.32)
	Between 50 and 100 m ²	3889 (46.37)
	More than 100 m ²	4025 (47.99)
	<i>in blank</i>	67 (0.80)

2. Experimental Design, Materials and Methods

2.1. Data collection

At the beginning of 2020, the world suffered a tremendous shock, caused by the health crisis of the SARS-CoV-2 virus. On March 11, the World Health Organization officially declared this

situation a pandemic. Three days later, the government of Spain established, through a Royal Decree, a state of alarm that came into force the following day, implementing a series of restrictive measures that had to be complied with [2]. One of these measures was strict home confinement, which was extended until the beginning of May when it was relaxed. The data described in this paper collect the responses from the Spanish population to a series of questions during this period. The valuable information provided explains several aspects related to the situation and the perception that the Spanish population had of the confinement and the state of alarm which for 99 days substantially restricted their freedom.

Between April 28 and May 14, 2020, the Research Group on Electoral Processes and Public Opinion of the University of Valencia (GIPEyOP) collected information from different social strata. The survey, organised into six blocks or sections, attained 8387 valid responses through a snowball sample design, initiated from a file of GIPEyOP collaborators (3236 at the time of the survey). GIPEyOP collaborators are people who selflessly participate with the research group by voluntarily answering and forwarding, at their convenience, the surveys generated by GIPEyOP. When we finish an investigation, a report is sent to them with the results obtained, in gratitude for their collaboration. If a person wishes to be part of this group of collaborators, they must fill in the form available on the group website <gipeyop.uv.es>. The link to this form is also available at the end of all our surveys to enrol more collaborators. Of course, a collaborator can unsubscribe at any time, via personal communication or by filling in another form available on GIPEyOP website.

The survey distribution process starts by sending by email a message to the GIPEyOP collaborators' list. Included in this message is a URL through which to access the online survey. They are asked to fill in the survey and to share it with their contacts. The forwarding of the survey is very simple to carry out, since the collaborators, in addition to completing the survey, can forward the received message to their contacts. But not only that. They can also share the URL with their acquaintances using social networks, with WhatsApp, Facebook, and Twitter being the most used. The survey has specific utilities to do that. In this way, starting from the initial list of collaborators, we managed to get the survey to a much larger segment of the population.

The URL that gave access to the survey was accompanied by the following message: "From the GIPEyOP research group of the University of Valencia we are studying the effects of the COVID19 crisis. We ask you for 10 min of your time to answer the survey and also that, please, share it with people over 17 around you. We appreciate that you disseminate the survey through social networks and among your contacts. The success of the research depends on you, and the variety and amount of information that we can collect. Thank you". In this way, the receiver of the link decided whether to access the questionnaire and/or resend it at that time, leave it for later or discard it definitively.

As mentioned above and can be inferred from the above explanation, a snowball, non-probabilistic method was used to select the sample. This technique does not guarantee the representativeness of the sample, among other issues, the sample obtained is partially conditioned to the place of work or residence of the person/people who initiate the process, as can be seen in Table 4. However, this procedure has some advantages over other sampling techniques: (i) it is an inexpensive and simple process, which has been described in some detail in the previous paragraph; (ii) it makes it possible to exploit the possibilities offered by new information technologies, mainly virtual social networks; (iii) it requires few human resources since interviewers are not necessary and the interviewed subjects themselves help to enrol new respondents; and (iv) makes it possible to sample populations that are difficult to access [3,4]. Furthermore, despite the biases in the data collected, when conditional inferences are made, the results of the modelling usually lead to conclusions equivalent to those obtained with representative samples [4,5].

To analyse the survey data, the individual responses obtained are weighted using post-stratification/calibration techniques to correct for biases in the collected sample [6]. To do that, we use two-class calibration approaches when we consider two variables to compute the sampling weights and marginal calibration (post-stratification) approaches when either one or more than two variables are employed. In our reports and models, we typically combine, in some

cases, two or more of the following variables: province of residence, habitat size, gender, age, and education level. In other cases, we use the combination of the variables R.VOTE (party voted in last elections) and province of residence. With these methods, we can compensate for the over-representation of some provinces or sociodemographic profiles in the sample.

Each of the questionnaires received was subjected to an intense filtering process to select only those questionnaires with minimum requirements in quality (internal consistency) and quantity of the available information. On the one hand, those questionnaires that did not contain a minimum number of responses were discarded. For example, as a rule, all samples that did not meet Section III were discarded. Questionnaires in which some sociodemographic variables, such as gender or province of residence, were not answered were also discarded. On the other hand, consistency tests were used, crossing pairs of variables, such as the size of the habitat and the province of residence. These actions led to 2636 responses being discarded. The validated dataset contains, as previously stated, a total of 8387 observations of 122 variables.

2.2. Questionnaire design and value of the data

As previously mentioned, the survey is structured in six sections or thematic blocks: (i) sociodemographic variables; (ii) residence during confinement; (iii) employment status; (iv) household chores; (v) health; and (vi) politics. The information collected in the first part of the study helps define the social and demographic profile of the respondents in the survey, information that is extremely relevant when analysing the results. The questions posed in the second section are aimed at assessing the conditions in which the population lived during that period and considering this to the respondent's perception of management of the crisis and health [7,8]. In addition, questions such as whether the home was the normal residence, its size, whether it had outdoor spaces, the number of people who lived together, and whether they were dependent can help draw conclusions about the consequences confinement has had for cohabitants, and how this could lead to a change in future habits, for example, valuing outdoor spaces more or even moving to rural areas [9]. The results of this survey, together with those of the research study [10], offer some of the key points that lead us to the conclusions reached in [11].

Section III of the survey is dedicated to the work/educational environment of the respondent to better understand how confinement has affected the performance of the population. The feelings that a person who works may have might be different from that of a student, a retiree, or someone unemployed with limited possibilities of finding work. In this sense, it is important to know how confinement affected studies or work performance, as well as to identify what new habits are likely to remain once this exceptional situation has ended. Some authors have already shown that women have seen their work performance affected by having to telework and complete household chores without having that spatial border between home and the workplace and that during confinement there has been no spatial or time delineation separating work from housework and the care of dependents [12–15].

The fourth part of the study focused on the tasks carried out by the respondent at home (before and during confinement). In this case, knowing how the fall of external services and formal and informal networks of care changed the usual way of distributing household chores and caring for minors or elderly dependents would indicate whether the gap between men and women has widened again, leading to a greater burden of work for women as they are the ones who usually assume the role of caregivers [1]. Some studies suggest that lack of mobility has had a greater impact on women than men [12,16–18]. This part of the study allows us to delve into the impact that the pandemic is having on the distribution of domestic tasks, and the widening of the gender gap.

The fifth section of the survey focused on the feelings and daily habits that the respondent had during confinement: fear of leaving the house, physical exercise, eating, sleep disorders, and symptoms that may indicate COVID-19 infection. Health care has been paramount throughout the pandemic. Numerous studies show that all the questions posed in this block are important

to the health of the immune system and, consequently, the ability to cope with this period in the best possible way [19].

In the last section of the survey, eight questions with a political profile were posed which tell us how the population feels about the way the health and economic crises have been managed by those in charge [20]. Respondents were asked which political party they voted for in the last elections, and what would be their choice if an election were held at the current time. By crossing these responses with the other variables, relations can be drawn, among other issues, between political ideology, sociodemographic variables, and perceptions related to work and conditions at homes.

Ethics Statement

At the beginning of the questionnaire, participants were informed that the survey was anonymous, voluntary, and confidential, as established in the current regulations on Personal Data Protection and Guarantee of Digital Rights. It was also indicated that the conclusions drawn from the survey would only be presented in aggregate form.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

CRedit Author Statement

Virgilio Pérez: Data curation, Investigation, Visualization, Writing – original draft, Writing – review & editing; **Cristina Aybar:** Data curation, Funding acquisition, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Jose M. Pavía:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Acknowledgments

The authors wish to thank an anonymous referee for its values comments and suggestions and all the GIPEyOP team, and mainly Alfredo Rubio and Rosa Roig, for their first-rate support and help on several phases of this research. Thanks are also due to Marie Hodkinson for translating the text of the paper into English and the SUPERA COVID-19 FUND, a joint tender offered by CRUE (Conference of Rectors of Spanish Universities), CSIC (the Spanish National Research Council), and Banco Santander (Santander Bank), for their financial support. The authors also acknowledge the support of Generalitat Valenciana through project AICO/2021/257 (Conselleria d'Innovació, Universitats, Ciència i Societat Digital).

Funding: This work was supported by the SUPERA COVID-19 FUND [Grant Number ROCOGIS (The Faces of Covid-19. Gender and Socioeconomic Impacts)].

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.dib.2021.107700](https://doi.org/10.1016/j.dib.2021.107700).

References

- [1] B. Larraz, R. Roig, C. Aybar, J.M. Pavía, COVID-19 and the gender division of housework. Traditional or new gender patterns?, (under review).
- [2] Ministerio de la Presidencia, Relaciones con las Cortes y Memoria Histórica, Real Decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el COVID-19, Boletín Oficial del Estado 67 sec. I (2020) 25390-25400. <https://www.boe.es/boe/dias/2020/03/14/pdfs/BOE-A-2020-3692.pdf>.
- [3] J. Kirchherr, K. Charles, Enhancing the sample diversity of snowball samples: recommendations from a research project on anti-dam movements in Southeast Asia, *PLoS ONE* 13 (2018) 1–17, doi:10.1371/journal.pone.0201710.
- [4] J.M. Pavía, I. Gil-Carceller, A. Rubio-Mataix, V. Coll, J.A. Alvarez-Jareño, C. Aybar, S. Carrasco-Arroyo, The formation of aggregate expectations: wisdom of the crowds or media influence? *Contemp. Soc. Sci.* 14 (2019) 132–143, doi:10.1080/21582041.2017.1367831.
- [5] J.M. Pavía, C. Aybar, Field rules and bias in random surveys with quota samples. An assessment of CIS surveys, *SORT* 42 (2018) 183–206, doi:10.2436/20.8080.02.74.
- [6] R. Valliant, J.A. Dever, F. Kreuter, *Practical Tools for Designing and Weighting Survey Samples*, Springer International Publishing AG, Cham, 2018.
- [7] M. Salvador-Ferrín, Transformación de la arquitectura desde el COVID-19, *Rev. Científica y Arbitr. Del Obs. Territ. Artes y Arquít.* 3 (2020) FINIBUS - ISSN 2737-6451 <https://publicacionescd.uleam.edu.ec/index.php/finibus/article/view/148>.
- [8] J.L. Nemecio, S. Tobón, Urbanización sustentable y resiliente ante el COVID-19: nuevos horizontes para la investigación de las ciudades, *Univ. y Soc.* 13 (2021) <https://rus.ucf.edu.cu/index.php/rus/article/view/1906>.
- [9] C.F. Sexto, Cidades dispersas e aldeias virtuais no pós-pandemia da COVID-19, *Finisterra* 55 (2021) 243–248, doi:10.18055/Finis20279.
- [10] V. Pérez, C. Aybar, J.M. Pavía, Dataset of the COVID-19 post lockdown survey conducted by GIPEyOP in Spain, (under review).
- [11] V. Pérez, C. Aybar, J.M. Pavía, COVID-19 and changes in social habits. Restaurant terraces, a booming space in cities. The case of Madrid, *Mathematics* 9 (2021), doi:10.3390/math9172133.
- [12] C. Collins, L.C. Landivar, L. Ruppanner, W.J. Scarborough, COVID-19 and the gender gap in work hours, *Gender Work Organ.* 28 (2021) 101–112, doi:10.1111/gwao.12506.
- [13] B.W. Cowan, Short-run effects of COVID-19 on U.S. worker transitions, *Natl. Bur. Econ. Res.* (2020) *Work. Pap. Ser. No. 27315*, doi:10.3386/w27315.
- [14] S. Schieman, P.J. Badawy, M.A. Milkie, A. Bierman, Work-life conflict during the COVID-19 pandemic, *Socius* 7 (2021) 2378023120982856, doi:10.1177/2378023120982856.
- [15] R. Roig, Teletrabajo y conciliación ante la igualdad de género. Una mirada desde las políticas públicas, in: A.I. Marades Puig (ed.), *Los Cuid. En La Era Covid-19 Análisis Jurídico, Económico y Político*, Tirant, 2021: pp. 299–314.
- [16] V.B. Catalá, La crisis de la COVID-19 y su impacto en los trabajos, *Anu. IET Trab. y Relac. Labor.* (2021) 187–209.
- [17] K. Hank, A. Steinbach, The virus changed everything, didn't it? Couples' division of housework and childcare before and during the Corona crisis, *J. Fam. Res.* (2021) 33, doi:10.20377/jfr-488.
- [18] C. Hupkau, C. Victoria, Covid-19 y desigualdad de género en España, *EsadeEcPol.* (2020). <https://dobetter.esade.edu/es/covid-19-desigualdad-genero>.
- [19] E. Andreu Cabrera, Actividad física y efectos psicológicos del confinamiento por COVID-19, *Rev. INFAD Psicol. Int. J. Dev. Educ. Psychol.* 2 (2020) 209–220, doi:10.17060/ijodaep.2020.n1.v2.1828.
- [20] M.H. Kim, W. Cho, H. Choi, J.Y. Hur, Assessing the South Korean model of emergency management during the COVID-19 pandemic, *Asian Stud. Rev.* 44 (2020) 567–578, doi:10.1080/10357823.2020.1779658.

Anexo V

Dataset of the COVID-19 post-lockdown survey conducted by GIPEyOP in Spain

Pérez, V., Aybar, C. & Pavía, J. M. (2022). *Data in Brief*, 40, 107763.

<https://doi.org/10.1016/j.dib.2021.107763>

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of the COVID-19 post-lockdown survey conducted by GIPEyOP in Spain

Virgilio Pérez^a, Cristina Aybar^a, Jose M. Pavía^{b,*}^a GIPEyOP, Department of Applied Economics, University of Valencia, Spain^b GIPEyOP, UMMICS, Department of Applied Economics, University of Valencia, Spain

ARTICLE INFO

Article history:

Received 8 October 2021

Revised 22 November 2021

Accepted 22 December 2021

Available online 24 December 2021

Keywords:

SARS-CoV-2

New normal

Change of habits

Social distancing

Online surveys



Snowball method

Political assessment

ABSTRACT

2020 was a year marked by COVID-19, an infectious disease caused by the SARS-CoV-2 virus. Since the official beginning of the pandemic (March 2020), the authorities in Spain have been imposing significant restrictions (mainly on mobility) to stop the spread of the disease. In October 2020, the research group GIPEyOP (Elections and Public Opinion Research Group from the University of Valencia) conducted a survey to analyse whether the Spanish population has maintained or modified their habits and customs once the strict measures imposed in Spain during the onset of the pandemic were relaxed. This article describes the dataset collected, which is provided as an attachment. The dataset is made up of 196 variables, following elimination of those variables that could potentially identify the respondents to ensure their anonymity. Over 22 days, from September 23 to October 14, 2020, GIPEyOP collected 1755 valid responses. Respondents were contacted by chain or snowball sampling via email and social media and answered a self-administered web questionnaire consisting of 40 questions. amongst other uses, the resulting dataset can be (re)used to analyse whether the period of home confinement that Spaniards experienced between March and June 2020 has caused them to change their habits and customs, such as how often they do sport or go to bars or restaurants. The data also permit the study of whether there have been changes in the distribution of

* Corresponding author.

E-mail address: pavia@uv.es (J.M. Pavía).Social media:  (V. Pérez),  (J.M. Pavía)

household chores by comparing three clearly differentiated moments (before confinement, during confinement and after confinement), what type of work (telework or face-to-face) the respondents would prefer or to know how the management of the crisis by govern authorities impacted on their votes preferences.

© 2021 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Social Science, Sociology, Political Science, Health, Economy
Specific subject area	Social Science (general), Public opinion, Political Science, Health, Economics
Type of data	Table (spreadsheet)
How the data were acquired	Data were collected using a self-administered online questionnaire. The questionnaire is provided in Spanish (original) and English (translated) as a supplementary file (word format). A snowball or chain sampling method was used to recruit respondents.
Data format	Raw
Description of data collection	The survey was carried out at the beginning of autumn 2020, once the state of alarm decreed by the Spanish Government had ended, a measure implemented to stop the evolution of the COVID-19 pandemic. The survey data was collected over twenty-two days (between September 23 and October 14, 2020).
Data source location	Country: Spain
Data accessibility	Data file (spreadsheet) is supplied as supplementary material with this article.

Value of the data

- This dataset offers information on various dimensions related to how the COVID-19 pandemic impacted on the Spanish population once a summer period with very relaxed measures had ended and a new COVID-19 wave was starting, with people having fresh in their memories a period of 99 days of home confinement just before summer.
- These data can be used by economists, sociologists, and political scholars to assess gender theories of behaviour within home in a scenario of the blurring of traditional gender roles.
- Economists could use these data to evaluate how the changes of habits have impacted on the economy.
- Political scientists could analyse how COVID-19 policies implemented by governments impacted on voting preferences.
- Psychologists may find these data useful to measure how the fear of being infected impacted on the preferences and habits of Spaniards when they spent time out of the house.
- This dataset is an example of how valuable information can be extracted from non-random samples.

1. Data Description

This document describes the data collected through a survey conducted between September 23 and October 14, 2020, a period in which SARS-CoV-2 coronavirus infections caused a second

wave of contagions in Spain. The survey data file spreadsheet that accompanies this article consists of 1755 rows and 196 columns. Each row presents an individual's response to the questionnaire and each column represents each of the variables generated from a questionnaire made up of 40 questions. The questionnaire, available as supplementary material to this article in Spanish and translated into English, is divided into 9 sections: (i) demographic variables, including age, gender, and province of residence, amongst others; (ii) daily life, focusing on the characteristics of the residence in which the respondent resides; (iii) employment situation, made up of 8 questions about work or study environment; (iv) household chores; (v) fears and cares, about the health conditions (physical and mental) of the respondents and about their habits after lockdown; (vi) holidays (destinations where respondents have chosen to go on holiday, and methods used to manage the corresponding booking arrangements); (vii) social life, including questions about respondents' social habits; (viii) digital divide, providing information about the apps that respondents use on their smartphones, or the type and quality of the Internet connection, amongst other questions; (ix) political management of the pandemic and related to the electoral debate.

The dataset and the dictionary of variables are supplied as supplementary material. In this dataset (spreadsheet) two types of missing values can be distinguished: blank cells, corresponding to non-response, and cells with the value N/A (Not Applicable), which refer to those questions not applicable for those surveyed for whom a certain question did not need answering due to their answers to previous questions (see [Table 1](#)). [Table 1](#) shows a brief description of the 196 variables available in the dataset. For reasons of space, the detail of the values for the PROV variable ([Section 1](#)), the Spanish province to which the respondent's municipality of residence belongs, is provided in [Table A1](#) (Appendix file) and not in [Table 1](#). Likewise, [Table A2](#) (Appendix file) offers the response options for the variables R.VOTE (vote recall) and VOTE (vote intention), both in [Section 9](#). This table contains the main political parties that stood in the 2019 general elections.

The mismatch between the number of questions (40) and the number of variables (196) comes from the fact that there are many questions for which more than a variable is extracted: (i) those questions in which the respondent could mark more than one answer have produced as many variables as there were answer options; (ii) questions in which the same topic was questioned at different moments in time; and (iii) the combination of both. This is the case of question 4001, about performing 14 household chores in 3 moments (before lockdown, during the lockdown, and after lockdown), giving rise to 42 variables (see [Table 2](#)). This question constitutes one of the central elements of the questionnaire and has helped, together with the data collected in [\[1\]](#), to carry out the research reported in [\[2\]](#) and [\[3\]](#).

[Table 3](#) provides information extracted from questions 7003 and 7004 ([Section 7](#)), referring to two specific customs and habits: (i) going to bars or restaurants and (ii) practising sports, before and after the state of alarm. The data provided in (i) was intensively exploited in [\[4\]](#) to study the impact of the pandemic on the Spanish restaurant sector. [Table 4](#) offers information extracted from question 8001 ([Section 8](#)), referring to the increase in the use of new technologies, both in the workplace and in the family or personal environment. This information allows us to establish relationships between how much and in what context Spaniards use technologies, and how this influences at (tele)work. [Table 5](#) presents a broad summary of the profile of the respondents in the survey. This explains the composition of the sample in terms of the main socio-economic-demographic characteristics, variables that in conjunction with R.VOTE (see [Table A2](#) in Appendix file) are routinely used to correct biases of surveys in political studies.

Table 1
Description of variables.

Section	Question code	Description	Values
		Respondent identification number	Number between 1 and 1755
		Time when the questionnaire was started / finished	Date and time
I	1001	Respondent's province of residence	See Table A1 (Appendix file)
	1002	Size of the municipality where the respondent resides	<ol style="list-style-type: none"> 1. Less than 2000 inhabitants 2. Between 2001 and 10,000 3. Between 10,001 and 50,000 4. Between 50,001 and 100,000 5. Between 100,001 and 400,000 6. Between 400,001 and 1,000,000 7. More than 1,000,000 inhabitants
	1003	Gender of the respondent	<ol style="list-style-type: none"> 1. Male 2. Female
	1004	Year of birthday	Number between 1919 and 2003
	1005	Highest education level achieved	<ol style="list-style-type: none"> 1. Without studies 2. Primary education 3. Secondary education 4. Job training 5. Baccaulaureate 6. University studies
II	2001	Do you live in the same home where you resided during the state of alarm (between 15 March and 21 June)?	<ol style="list-style-type: none"> 1. Yes, it is my usual home. 2. Yes, although it is not my usual residence. 3. Yes, I would like to change, but I have no choice. 4. No, I have moved to another residence. 5. No, I have returned to my usual residence.
	2002	How many people do you live with?	<ol style="list-style-type: none"> 1. None 2. One 3. Two 4. Three 5. Four 6. Five or more
	2002A	Could you please indicate the number of dependents you live with? <i>Note: only to be answered by respondents who did not choose 1 in question 2002.</i>	<ol style="list-style-type: none"> 1. None 2. One 3. Two 4. Three 5. Four 6. Five or more
	2002B	Could you please indicate the number of high-risk people you live with? <i>Note: only to be answered by respondents who did not choose 1 in question 2002.</i>	<ol style="list-style-type: none"> 1. None 2. One 3. Two 4. Three 5. Four 6. Five or more

(continued on next page)

Table 1 (continued)

Section	Question code	Description	Values
III	3001	Comparison of the current employment situation with respect to that experienced before / during the state of alarm.	<ol style="list-style-type: none"> 1. Same 2. Better 3. Worse
	3002	Respondent's employment situation.	<ol style="list-style-type: none"> 1. I am a salaried employee and telework. 2. I am a salaried employee and I work outside the home. 3. I am a salaried employee and I combine telework and work outside the home. 4. I am self-employed and telework. 5. I am self-employed and I work outside the home. 6. I am self-employed and I combine telework and work outside the home. 7. I am self-employed with no possibility of practising my profession. 8. I am on an ERTE (Temporary Lay-off Plan). 9. I was fired after the state of alarm period. 10. I was fired during the state of alarm period. 11. Sick leave/pregnancy. 12. I am unemployed or on leave of absence. 13. Retired 14. Student 15. Unpaid work at home. 16. I work without a contract outside the home. 17. Other
	3003	Respondent's labour sector. <i>Note: only to be answered by respondents who chose 1, 2 or 3 in question 3002.</i>	<ol style="list-style-type: none"> 1. Private sector 2. Public sector 3. Both sectors
	3003A1	Do you feel that your productivity at work has been affected since the end of the state of alarm? <i>Note: only to be answered by respondents who chose 1 to 6 in question 3002.</i>	<ol style="list-style-type: none"> 1. Yes, I have a higher performance. 2. Yes, I have a lower performance. 3. No.
	3003A2	Do you think your work is threatened by this second wave of infections? <i>Note 1: only to be answered by respondents who chose 1 to 6 or 16 in question 3002.</i> <i>Note 2: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. Yes, because of a lack of economic activity due to the crisis. 2. Yes, due to staff cuts. 3. Yes, due to salary cuts. 4. Yes, because of having had to help in the family environment and having underperformed at work. 5. No, everything will stay the same.

(continued on next page)

Table 1 (continued)

Section	Question code	Description	Values
	3002B	What is your experience of working at home after the end of the state of alarm? <i>Note 1: only to be answered by respondents who chose 1, 3, 4 or 6 in question 3002.</i> <i>Note 2: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. I am making better use of my time than in my place of work. 2. It is difficult to reconcile work and family life. 3. I wouldn't mind continuing to telework. 4. I prefer to commute to my place of work. 5. I would like to alternate between the two options. 6. I was already teleworking before the state of alarm.
	3002C1	Concern about the operation of the school year. <i>Note: only to be answered by respondents who chose 14 in question 3002.</i>	<ol style="list-style-type: none"> 1. Yes, I am afraid of being in the classroom and catching the virus. 2. Yes, I am not sure how the classes will be taught. 3. Yes, I have not received enough information. 4. Yes, I do not have the conditions to study from home if necessary. 5. Yes, for other reasons. 6. I have no concerns.
	3002C2	How are you going to cover the costs of university studies? <i>Note: only to be answered by respondents who chose 14 in question 3002.</i>	<ol style="list-style-type: none"> 1. My family. 2. I have started to work since the end of the state of alarm. 3. I was already working before the state of alarm started. 4. I have savings. 5. A scholarship. 6. I have applied for a scholarship due to lack of resources. 7. Others.
IV	4001	Weekly frequency of the corresponding task (14 tasks and 3 moments in time). <i>Note: tasks and moments in time are shown in Table 2.</i>	0 times to 7 times; No proceed to response
	4002	Outside help with housework before / after lockdown (weekly frequency).	0 times to 7 times; No proceed to response
V	5001	Fear of leaving the home	<ol style="list-style-type: none"> 1. I have practically not gone out, but I am not afraid. 2. I have practically not gone out because I am afraid. 3. I go out alone to carry out basic tasks (walking the dog, shopping, work, care ...), although I am afraid. 4. I go out alone to carry out basic tasks (walking the dog, shopping, work, care ...), and I am not afraid. 5. I go out normally and I am not afraid. 6. I go out normally, but I am a bit afraid.

(continued on next page)

Table 1 (continued)

Section	Question code	Description	Values
	5002	How well did you sleep compared to before the lockdown?	<ol style="list-style-type: none"> 1. Same 2. Better 3. Worse
	5003	<p>What hygiene measures do you usually take?</p> <p><i>Note: you can indicate more than one option.</i></p>	<ol style="list-style-type: none"> 1. I change my masque after its useful life. 2. I wear a masque, but I reuse it more than I should. 3. When I take off my masque, I am careful where I keep it. 4. I sanitise my hands whenever I touch something (public transport, coins...). 5. I take care of my immune system (food, vitamin supplements, physical exercise...). 6. I do not follow any special measures; I wear a masque out of obligation.
VI	6001	<p>What did you do during your holidays last summer (2019)?</p> <p><i>Note: you can indicate more than one option.</i></p>	<ol style="list-style-type: none"> 1. Worked. I did not have any holiday time. 2. Worked and enjoyed a few days of holiday. 3. Did not travel due to lack of financial resources. 4. Went to the countryside, to my second residence. 5. Went to the beach, to my second residence. 6. Went to the mountains, to my second residence. 7. Went on a trip in Spain. 8. Went on a trip outside of Spain. 9. Others.
	6002	How much time have you had for holidays since the state of alarm ended?	<ol style="list-style-type: none"> 1. I have had no holiday time. 2. Less than a week. 3. About fifteen days. 4. Between fifteen days and a month. 5. More than one month.
	6002A	<p>What did you do during your holidays this summer (2020)?</p> <p><i>Note 1: only to be answered by respondents who did not choose 1 in question 6002.</i></p> <p><i>Note 2: you can indicate more than one option.</i></p>	<ol style="list-style-type: none"> 1. Did not travel due to lack of financial resources. 2. Did not travel due to economic uncertainty. 3. Did not travel due to fear of contagion. 4. Went to the countryside, to my second residence. 5. Went to the beach, to my second residence. 6. Went to the mountains, to my second residence. 7. Went on a trip in Spain. 8. Went on a trip outside of Spain. 9. Others.
	6002B	<p>How have you managed to make arrangements for your holidays?</p> <p><i>Note: only to be answered by respondents who chose 4 to 8 in question 6002A.</i></p>	<ol style="list-style-type: none"> 1. I have not needed to make any arrangements. 2. By phone. 3. By Internet.

(continued on next page)

Table 1 (continued)

Section	Question code	Description	Values
	6002C	Were you afraid of COVID-19 on your holiday? <i>Note 1: only to be answered by respondents who don't chose 4 to 8 in question 6002A.</i> <i>Note 2: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. Yes, people did not wear masks. 2. Yes, people did not keep a safe distance. 3. Yes, there was an outbreak in a nearby location. 4. Yes, in my local environment there was a contagion. 5. Yes, around me were high risk people. 6. I was afraid out of respect for the disease. 7. I was not afraid.
VII	7001	Have you changed your habits regarding the number of people you interact with? <i>Note: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. No, same as before. 2. Yes, I only socialise with the people I live with. 3. Yes, I only socialise with the people within my closest circle. 4. Yes, I have reduced the number of people I interact with.
	7002	Have you changed your habits outside your usual residence?	<ol style="list-style-type: none"> 1. I only choose open spaces. 2. I might choose closed spaces as long as there is good ventilation. 3. I usually avoid closed spaces. 4. I do not restrict my choices.
	7003	How often did you used to go to a bar or restaurant before the state of alarm / since the state of alarm ended?	<ol style="list-style-type: none"> 1. Every day 2. 5 or 6 days a week 3. 3 or 4 days a week 4. 1 or 2 days a week 5. Never
	7004	How often did you play sports before the state of alarm / since the state of alarm ended?	<ol style="list-style-type: none"> 1. Every day 2. 5 or 6 days a week 3. 3 or 4 days a week 4. 1 or 2 days a week 5. Never
VIII	8001	Do you use a smartphone or the internet more? (for work / for personal issues) <i>Note: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. No, same as before 2. Yes, more video conferencing 3. Yes, more emails 4. Yes, more use of Twitter 5. Yes, more use of Facebook 6. Yes, more use of WhatsApp 7. Yes, more use of Telegram 8. Yes, more use of Instagram
	8002	Select the apps you used before the pandemic / currently use. <i>Note: you can indicate more than one option.</i>	<ol style="list-style-type: none"> 1. Facebook 2. Instagram 3. WhatsApp 4. Telegram 5. Twitter 6. YouTube 7. E-mail (for personal use) 8. E-mail (to work) 9. Skype 10. Zoom 11. Teams 12. TikTok 13. Tinder 14. Others

(continued on next page)

Table 1 (continued)

Section	Question code	Description	Values
	8003	What device do you usually use to connect to the internet?	<ol style="list-style-type: none"> 1. Smartphone 2. Tablet 3. Laptop 4. PC 5. Smartwatch 6. Others
	8004	How do you prefer to communicate when you do not do so in person? (for work / for personal issues)	<ol style="list-style-type: none"> 1. Telephone call using a landline 2. Telephone call using a mobile phone 3. Message written by WhatsApp or Telegram 4. Written email 5. Audio message by WhatsApp or Telegram 6. Videoconference (WhatsApp, Skype, Zoom ...)
	8005	Do you usually have problems with your internet connection?	<ol style="list-style-type: none"> 1. No. 2. Yes, the network goes down sometimes 3. Yes, the network crashes 4. Yes, the network is slow
IX	9001	Valuation of the local/regional/national government in terms of management of the health/economic crisis (before summer/second wave).	0 (very bad) to 10 (very good)
	9002		
	9003	Could you tell me which party you voted for in the last General Election?	<ol style="list-style-type: none"> 1. List of parties in Table A2 (Appendix file) 2. Others 3. I was not old enough to vote 4. Abstention 5. I voted blank
	9004	If a congressional election were held today, which party would you vote for?	<ol style="list-style-type: none"> 1. List of parties in Table A2 (Appendix file) 2. Others 3. I would not vote 4. I don't have the right to vote
		Time taken to complete the questionnaire	Numbers of seconds taken
	Time needed to complete Sections (I–IX)	Numbers of seconds taken	

Table 2

Weekly frequency (average number of days) of the respondent carrying out certain household chores before lockdown (BL), during lockdown (DL), and after lockdown (AL), by gender.

Household chores	Male			Female		
	BL	DL	AL	BL	DL	AL
Preparing midday meal	3.28	4.30	3.59	4.26	5.56	4.71
Cleaning the bathroom	1.37	1.92	1.36	2.57	3.59	2.79
Helping with children's homework	1.10	1.63	1.03	1.15	1.94	1.18
Playing with minors	1.84	2.05	1.77	1.97	2.36	2.01
Preparing the dinner	4.43	4.83	4.47	5.19	5.59	5.21
Bathing dependant persons	0.62	0.69	0.64	1.09	1.19	1.09
Leaving the house to look after other dependents	0.61	0.67	0.58	0.93	1.06	0.89
Washing up after meals	4.22	4.57	4.13	4.90	5.40	5.03
Dusting	1.34	1.79	1.27	2.05	2.88	2.15
Cleaning the floor	1.67	2.21	1.71	2.49	3.60	2.69
Going out for grocery shopping	2.65	2.26	2.54	2.56	1.48	2.18
Doing the washing	1.70	1.80	1.69	3.01	3.13	3.14
Ironing	0.66	0.63	0.61	1.17	1.14	1.18
Throwing out the rubbish	3.96	3.97	3.93	3.80	3.75	3.78

Table 3

Comparison of habits and customs (before and after the state of alarm).

Category	Bars/Restaurants		Sport	
	Before	After	Before	After
Every day	116 (7.92)	39 (2.75)	295 (20.04)	230 (15.71)
5 or 6 days a week	119 (8.12)	28 (1.97)	231 (15.69)	188 (12.84)
3 or 4 days a week	340 (23.21)	111 (7.82)	427 (29.01)	348 (23.77)
1 or 2 days a week	785 (53.58)	729 (51.37)	343 (23.30)	472 (32.24)
Never	105 (7.17)	512 (36.08)	176 (11.96)	226 (15.44)

Table 4

Comparison of frequency of use of technological means (for work and for personal issues).

Category	For work	For personal issues
No, same as before	404 (23.02)	570 (32.48)
Yes, more video conferencing	563 (32.08)	508 (28.95)
Yes, more emails	431 (24.56)	184 (10.48)
Yes, more use of Twitter	43 (2.45)	195 (11.11)
Yes, more use of Facebook	40 (2.28)	238 (13.56)
Yes, more use of WhatsApp	324 (18.46)	707 (40.28)
Yes, more use of Telegram	40 (2.28)	118 (6.72)
Yes, more use of Instagram	34 (1.94)	231 (13.16)

Table 5

Respondent characteristics (n = 1755).

Characteristics	Category	Frequency (%)
Gender	Male	955 (54.42)
	Female	743 (42.34)
	<i>in blank</i>	57 (3.25)
Age (years)	< 20	6 (0.34)
	20–25	79 (4.50)
	26–30	117 (6.67)
	31–35	105 (5.98)
	36–40	144 (8.21)
	41–45	175 (9.97)
	46–50	184 (10.48)
	51–55	190 (10.83)
	56–60	220 (12.54)
	61–65	201 (11.45)
	66–70	149 (8.49)
	>70	151 (8.60)
	<i>in blank</i>	34 (1.94)
Employment situation	I am a salaried employee and telework.	137 (7.81)
	I am a salaried employee and I work outside the home.	538 (30.66)
	I am a salaried employee and I combine telework and work outside the home.	204 (11.62)
	I am self-employed and telework.	22 (1.25)
	I am self-employed and I work outside the home.	68 (3.87)
	I am self-employed and I combine telework with work outside the home.	42 (2.39)
	I am self-employed with no possibility of practising my profession.	9 (0.51)
	I am on an ERTE.	19 (1.08)
	I was fired after the state of alarm period.	18 (1.03)
	I was fired during the state of alarm period.	14 (0.80)
	Sick leave/pregnancy	19 (1.08)
	I am unemployed or on leave of absence.	97 (5.53)
	Retired	371 (21.14)
	Student	52 (2.96)
	Unpaid work at home	21 (1.20)
	I work without a contract outside my home.	10 (0.57)
Other	49 (2.79)	
<i>in blank</i>	65 (3.70)	
Education	Without studies	1 (0.06)
	Primary education	46 (2.62)
	Secondary education	63 (3.59)
	Job training	147 (8.38)
	Baccalaureate	187 (10.66)
	University studies	1273 (72.54)
<i>in blank</i>	38 (2.17)	
Residence municipality size (inhabitants)	Less than 2000 inhabitants	57 (3.25)
	Between 2001 and 10,000	192 (10.94)
	Between 10,001 and 50,000	379 (21.60)
	Between 50,001 and 100,000	189 (10.77)
	Between 100,001 and 400,000	225 (12.82)
	Between 400,001 and 1,000,000	352 (20.06)
	More than 1,000,000 de inhabitants	290 (16.52)
<i>in blank</i>	71 (4.05)	

2. Experimental Design, Materials and Methods

2.1. Data collection

In early 2020, the world was exposed to an extreme threat, that of the SARS-CoV-2 coronavirus health crisis. On 14 March, the government of Spain imposed, through a Royal Decree, a series of mandatory restrictive measures [5]. One of these measures was strict home confinement, which lasted 99 days. The favourable evolution of the pandemic allowed the authorities to lift the restrictions previously imposed from the end of spring, with life returning to what is generally referred to as the new normal [6], a situation in which the strict measures imposed at the beginning of the pandemic have been relaxed, but some limitations remain in place. However, the impact of the health crisis on Spanish society has led to a substantial change in habits and customs, for example, in the way people relate to each other or how public spaces are used and their frequency of use, in particular the use of closed spaces, such as bars and restaurants [4]. The data described in the present study show the responses of the Spanish population to a series of contextualized questions in a new period marked by the end of home confinement (which lasts more than three months) and after a summer in which the measures to contain the spread of the virus had been relaxed; just in a moment when Spain was seeing to have the highest number of people infected by COVID-19 of any European country [7].

Between September 23 and October 14, 2020, the Research Group on Electoral Processes and Public Opinion of the University of Valencia (GIPEyOP) prepared a survey to collect information from different social strata related to this context. The information gathered enables the comparison of the valuation and perception of Spaniards at different times of the pandemic and on topics as varied as sports practice, the use of mobile devices and the internet, evaluation of government officials and their management of the health and economic crisis, or the involvement of different members of the family unit in housework.

The survey, organised into nine blocks or sections, collected 1755 valid responses through a snowball sample design, initiated from a file of GIPEyOP collaborators. GIPEyOP collaborators are people who selflessly participate with the research group by voluntarily answering and forwarding, at their convenience, the surveys generated by GIPEyOP. When we finish an investigation, a report is sent to them with the results obtained, in gratitude for their collaboration. If a person wishes to be part of this group of collaborators, they must fill in the form available on the group website <gipeyop.uv.es>. The link to this form is also available at the end of all our surveys to enrol more collaborators. Of course, a collaborator can unsubscribe at any time, via personal communication or by filling in another form available on the GIPEyOP website.

The survey distribution process starts by sending by email a message to the GIPEyOP collaborators' list. Included in this message is a URL through which to access the online survey. They are asked to fill in the survey and to share it with their contacts. The forwarding of the survey is very simple to carry out, since the collaborators, in addition to completing the survey, can forward the received message to their contacts. But not only that. They can also share the URL with their acquaintances using social networks, with WhatsApp, Facebook, and Twitter being the most used. The survey has specific utilities to do that. In this way, starting from the initial list of collaborators, we managed to get the survey to a much larger segment of the population.

The URL that gave access to the survey was accompanied by the following message: "From the GIPEyOP research group of the University of Valencia we are studying the effects of the COVID19 crisis. We ask you for 10 min of your time to answer the survey and also that, please, share it with people over 17 in your environment. We appreciate that you disseminate the survey through social networks and amongst your contacts. The success of the research depends on you, and the variety and amount of information that we can collect. Thank you". In this way, the receiver of the link decided whether to access the questionnaire and/or resend it at that time, leave it for later or discard it definitively.

As mentioned above, the snowball technique was used to select the sample. This non-probabilistic method does not guarantee the representativeness of the sample, amongst other

issues, the sample obtained is partially conditioned to the place of work or residence of the person/people who initiate the process. However, this procedure has some advantages over other sampling techniques: (i) it is an inexpensive and simple process, which has been described in some detail in the previous paragraph; (ii) it makes it possible to exploit the possibilities offered by new information technologies, mainly virtual social networks; (iii) it requires few human resources since interviewers are not necessary and the interviewed subjects themselves help to enrol new respondents; (iv) makes it possible to sample populations that are difficult to access [8,9]. But not only that. This technique allows, unlike other questionnaires that are repeated over time, to compare opinions, habits, and feelings of roughly similar groups at different moments in time. Furthermore, despite the biases in the data collected, when conditional inferences are made, the results of the modelling usually lead to conclusions equivalent to those obtained with representative samples [10].

To analyse the survey data, the individual responses obtained are weighted using post-stratification/calibration techniques to correct for biases in the collected sample [11]. To do that, we use two-class calibration approaches when we consider two variables to compute the sampling weights and marginal calibration (post-stratification) approaches when either one or more than two variables are employed. In our reports and models, we typically combine, in some cases, two or more of the following variables: province of residence, habitat size, gender, age, and education level. In other cases, we use the combination of the variables R.VOTE (party voted in last elections) and the province of residence. With these methods, we can compensate for the over-representation of some provinces or sociodemographic profiles in the sample.

Each of the questionnaires received was subjected to an intense filtering process to select only those questionnaires with minimum requirements in quality (internal consistency) and quantity of the available information. On the one hand, those questionnaires that did not contain a minimum number of responses were discarded. For example, as a rule, all samples that did not meet Section 3 were discarded. On the other hand, consistency tests were used, crossing pairs of variables, such as the size of the habitat and the province of residence. These actions resulted in the elimination of 233 responses (11.7% of the total). The validated dataset contains, as indicated above, a total of 1755 observations in 196 variables.

2.2. Value of the data

As previously mentioned, the survey is structured in nine sections or thematic blocks. The information collected in the first part of the study helps to define the social and demographic profile of the respondents in the survey, information that is of relevance when analysing the results. The questions in the second section (daily life) help assess, amongst other issues, whether Spaniards have modified their habits in terms of caring for dependant people or people at high risk of contagion, particularly important considering the high mortality that SARS-CoV-2 caused in the elderly [12]. Section 3 of the survey is dedicated to the respondent's work/educational environment and helps identify how, and to what extent, the period of confinement has modified how Spaniards work/study, and whether it has prompted a transition period towards a new paradigm marked by teleworking and online education [13,14]. In both Sections 2 and 3 some questions enable the perception of the respondent at the time the survey was conducted (October 2020) to be compared with their perception during the period of confinement (second quarter of 2020, when a further survey was conducted [1]). The feelings that a person who works may have might be different from that of a student, a retiree, or someone unemployed with limited possibilities of finding work. In this sense, it is important to know how confinement affected studies or work performance, as well as to identify what new habits and customs are likely to remain once this exceptional situation has ended.

Section 4 focuses on analysing the completion of certain household chores (before, during, and after confinement). This block of questions reveals whether the situation of forced cohabitation changed the usual way of distributing chores related to the home and the care of minors or dependant elderly and whether, once the confinement period ended, these possible changes

have been maintained or, on the contrary, things have returned to the way they were, or perhaps they depend on some co-variables, such as the habitat size [3] or the modality of work (telework or face-to-face) [15].

Section 5 of the survey focuses on three aspects related to the respondent's health: (i) fear of leaving the house; (ii) sleep disorders; and (iii) hygienic-sanitary measures. Studies have been published throughout the pandemic that demonstrates the relationship between the issues raised in this block and protection against the coronavirus [16]. Knowing whether these possible disorders have disappeared since the end of the state of alarm or whether they have been reduced may be important when considering the need for the implementation of therapeutic measures and for establishing certain action protocols.

Section 6 is orientated towards analysing decisions that the respondents made regarding their summer holidays of 2019 (before the pandemic) and 2020 (after the lockdown). The questions posed provide information on how the customs and habits of Spanish society have changed in this regard, making these data especially interesting for implementing new business models and for all agents related to the tourism sector, a sector severely affected by the pandemic [17].

Section 7 of the questionnaire includes questions related to certain social and consumption customs and habits, such as the frequency with which the respondents practised/practise sports or the regularity with which they frequented/frequent bars and restaurants (before and after the state of alarm). These types of questions help identify whether the behaviour patterns of Spaniards have been affected by the extreme situation experienced in 2020 and whether these possible changes will lead to transformations that may require certain sectors to adapt, such as the hospitality sector is doing, to guarantee their survival. In Madrid, for example, the limitations on mobility for so many months and the fear of contagion in closed spaces have meant that, with the support of the local government, bars and restaurants have had to reinvent themselves, adding/expanding outdoor terraces to adapt to the new demands of their customers [4].

Section 8 is dedicated to the use of the internet and applications on mobile devices, both in the workplace and in personal time. These questions help to identify whether the way Spaniards used to communicate has changed because of confinement and whether the technological resources currently available are adequate and sufficient. The emergence of teleworking has led to a significant increase in cyberattacks, both in the public and private sectors [18], posing major business challenges while generating new jobs and business opportunities.

In the last section of the survey, four political profile questions are asked that give an insight into how the population values how government officials have managed the health and economic crisis [19] and how this impact on their vote preferences. Specifically, it assesses how effectively the respondents think the government has managed the situation at local, regional, and national levels, comparing two moments in time: before the state of alarm was decreed and in the moment of conducting the survey, just when the second wave of infections began in Spain. Respondents are also asked which political party they voted for in the last election and what their choice would be if elections were held at present. By crossing these responses with the other variables, relationships can be established between political ideology, sociodemographic variables, and perceptions related to work and household conditions.

Ethics Statements

At the beginning of the questionnaire, participants were informed that the survey was anonymous, voluntary and confidential, as established in the current regulations on Personal Data Protection and guarantee of digital rights. It was also indicated that the conclusions drawn from the survey would only be presented in aggregate form.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

CRedit Author Statement

Virgilio Pérez: Data curation, Investigation, Visualization, Writing – original draft, Writing – review & editing; **Cristina Aybar:** Data curation, Funding acquisition, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Jose M. Pavía:** Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

This work was supported by the SUPERA COVID-19 FUND [Grant No. ROCOGIS (The Faces of COVID-19. Gender and Socioeconomic Impacts)] and by Conselleria d'Innovació, Universitats, Ciència i Societat Digital [Grant No. AICO/2021/257].

Acknowledgements

The authors wish to thank all the GIPEyOP team, and mainly Alfredo Rubio and Rosa Roig, for their first-rate support and help on several phases of this research. Thanks are also due to Marie Hodgkinson for translating the text of the paper into English and the SUPERA COVID-19 FUND, a joint tender offered by CRUE (Conference of Rectors of Spanish Universities), CSIC (the Spanish National Research Council) and Banco Santander (Santander Bank), for their financial support. The authors also acknowledge the support of Generalitat Valenciana through project AICO/2021/257 (Conselleria d'Innovació, Universitats, Ciència i Societat Digital).

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.dib.2021.107763](https://doi.org/10.1016/j.dib.2021.107763).

References

- [1] V. Pérez, C. Aybar, J.M. Pavía, Dataset of the COVID-19 lockdown survey conducted by GIPEyOP in Spain, *Data in Brief* 40 (2022) 107700, doi:[10.1016/j.dib.2021.107700](https://doi.org/10.1016/j.dib.2021.107700).
- [2] B. Larraz, R. Roig, C. Aybar, J.M. Pavía, COVID-19 and the gender division of housework. Traditional or new gender patterns? (Submitted for publication).
- [3] R. Roig, C. Aybar, J.M. Pavía, COVID-19, gender housework division and habitat size in Spain, (Submitted for publication).
- [4] V. Pérez, C. Aybar, J.M. Pavía, COVID-19 and changes in social habits. Restaurant terraces, a booming space in cities. The case of Madrid, *Mathematics*. 9 (2021) 2133, doi:[10.3390/math9172133](https://doi.org/10.3390/math9172133).
- [5] Ministerio de la Presidencia, Real Decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el COVID-19, *Boletín Oficial del Estado* 67 (sec. I) (2020) 25390–25400 <https://www.boe.es/boe/dias/2020/03/14/pdfs/BOE-A-2020-3692.pdf>.
- [6] Ministerio de la Presidencia, Nueva normalidad, La Moncloa (2020). <https://www.lamoncloa.gob.es/covid-19/Paginas/nueva-normalidad.aspx> (accessed November 20, 2021).
- [7] DNS, Coronavirus (COVID-19) - 23 de octubre 2020, Departamento de Seguridad Nacional (2020). <https://www.dsn.gob.es/es/actualidad/sala-prensa/coronavirus-covid-19-23-octubre-2020> (accessed November 20, 2021).
- [8] J. Kirchherr, K. Charles, Enhancing the sample diversity of snowball samples: recommendations from a research project on anti-dam movements in Southeast Asia, *PLoS ONE* 13 (2018) 1–17, doi:[10.1371/journal.pone.0201710](https://doi.org/10.1371/journal.pone.0201710).
- [9] J.M. Pavía, I. Gil-Carceller, A. Rubio-Mataix, V. Coll, J.A. Alvarez-Jareño, C. Aybar, S. Carrasco-Arroyo, The formation of aggregate expectations: wisdom of the crowds or media influence? *Contemp. Soc. Sci.* 14 (2019) 132–143, doi:[10.1080/21582041.2017.1367831](https://doi.org/10.1080/21582041.2017.1367831).
- [10] J.M. Pavía, C. Aybar, Field rules and bias in random surveys with quota samples. An assessment of CIS surveys, *SORT* 42 (2018) 183–206, doi:[10.2436/20.8080.02.74](https://doi.org/10.2436/20.8080.02.74).
- [11] R. Valliant, J.A. Dever, F. Kreuter, *Practical Tools for Designing and Weighting Survey Samples*, Springer International Publishing AG, Cham, 2018.

- [12] C. Bonanad, S. García-Blas, F. Tarazona-Santabalbina, J. Sanchis, V. Bertomeu-González, L. Fácila, A. Ariza, J. Núñez, A. Cordero, The effect of age on mortality in patients with COVID-19: a meta-analysis with 611,583 subjects, *J. Am. Med. Dir. Assoc.* 21 (2020) 915–918, doi:[10.1016/j.jamda.2020.05.045](https://doi.org/10.1016/j.jamda.2020.05.045).
- [13] P. Cay-Larcuen, H. Bellido-Bellón, Análisis demográfico del teletrabajo en España entre los años 2019-2021, ZAGUAN, Repositorio Institucional de Documentos, 2021 <https://zaguan.unizar.es/record/106964>.
- [14] L. García-Aretio, COVID-19 y educación a distancia digital: preconfinamiento, confinamiento y posconfinamiento, *RIED. Rev. Iberoam. Educ. a Distancia.* 24 (2020) 9–32, doi:[10.5944/ried.24.1.28080](https://doi.org/10.5944/ried.24.1.28080).
- [15] R. Roig, C. Pineda, El teletrabajo y la conciliación: dos políticas públicas diferentes, *GIGAPP Estud. Work. Pap* 7 (2020) 593–608 <http://www.gigapp.org/ewp/index.php/GIGAPP-EWP/article/view/203>.
- [16] E. Andreu-Cabrera, Actividad física y efectos psicológicos del confinamiento por COVID-19, *Rev. INFAD Psicol. Int. J. Dev. Educ. Psychol.* 2 (2020) 209–220, doi:[10.17060/ijodaep.2020.n1.v2.1828](https://doi.org/10.17060/ijodaep.2020.n1.v2.1828).
- [17] G. García-Andrés, A. García-Baquero, El golpe al turismo y la recuperación de la economía española, *Cuad. Inf. Económica.* 277 (2020) 25–32 <https://www.funcas.es/wp-content/uploads/2020/08/CIE277art04.pdf>.
- [18] M. Jurado, La Ciberseguridad en El Marco Europeo, El caso de España, Universidad de Almería, 2020 <http://repositorio.ual.es/handle/10835/9544>.
- [19] M.H. Kim, W. Cho, H. Choi, J.Y. Hur, Assessing the South Korean model of emergency management during the COVID-19 pandemic, *Asian Stud. Rev.* 44 (2020) 567–578, doi:[10.1080/10357823.2020.1779658](https://doi.org/10.1080/10357823.2020.1779658).

Anexo VI


COVID-19 and Changes in Social Habits. Restaurant Terraces, a Booming Space in Cities. The Case of Madrid

Pérez, V., Aybar, C. & Pavía, J. M. (2021). *Mathematics*, 9(17), 2133.

<https://doi.org/10.3390/math9172133>

Article

COVID-19 and Changes in Social Habits. Restaurant Terraces, a Booming Space in Cities. The Case of Madrid

Virgilio Pérez ¹, Cristina Aybar ¹ and Jose M. Pavía ^{2,*}

¹ GIPEyOP, Área de Métodos Cuantitativos, Departamento de Economía Aplicada, Facultat d'Economía, Universitat de Valencia, 46020 Valencia, Spain; virgilio.perez@uv.es (V.P.); cristina.aybar@uv.es (C.A.)

² GIPEyOP, UMICCS, Área de Métodos Cuantitativos, Departamento de Economía Aplicada, Facultat d'Economía, Universitat de Valencia, 46020 Valencia, Spain

* Correspondence: pavia@uv.es

Abstract: The COVID-19 pandemic and the fear experienced by some of the population, along with the lack of mobility due to the restrictions imposed, has modified the social behaviour of Spaniards. This has had a significant effect on the hospitality sector, viewed as being an economic and social driver in Spain. From the analysis of data collected in two of our own non-probabilistic surveys (N ~ 8400 and N ~ 2000), we show how, during the first six months of the pandemic, Spaniards notably reduced their consumption in bars and restaurants, also preferring outdoor spaces to spaces inside. The restaurant sector has needed to adapt to this situation and, with the support of the authorities (regional and local governments), new terraces have been allowed on pavements and public parking spaces, modifying the appearance of the streets of main towns and cities. This study, focused on the city of Madrid, analyses the singular causes that have prompted this significant impact on this particular city, albeit with an uneven spatial distribution. It seems likely that the new measures will leave their mark and some of the changes will remain. The positive response to these changes from the residents of Madrid has ensured the issue is being widely debated in the public arena.



Citation: Pérez, V.; Aybar, C.; Pavía, J.M. COVID-19 and Changes in Social Habits. Restaurant Terraces, a Booming Space in Cities. The Case of Madrid. *Mathematics* **2021**, *9*, 2133. <https://doi.org/10.3390/math9172133>

Academic Editors: María Del Mar Rueda and Andrés Cabrera-León

Received: 14 August 2021
Accepted: 31 August 2021
Published: 2 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: COVID-19; lockdown; restaurant sector; web survey; snowball survey

1. Introduction

On 11 March 2020, COVID-19 was officially declared a pandemic by the World Health Organisation (WHO). Two days later, the Government of Spain, through a Royal Decree, declared a state of alarm for the entire country, which came into force the following day [1]. In order to curb the transmission and spread of the virus, Spain decreed mandatory home confinement for residents throughout the country, excluding essential workers (in the food, health or safety sectors). The confinement was extended on successive occasions, finally ending on 21 June 2020, 103 days later. Spain implemented one of the strictest confinements worldwide. Specifically, from the analysis of 23 indicators, Spain scored between 80 and 90 points on a scale of 100 [2]. During the first 52 days of confinement, people were only allowed to leave their homes to buy food or essential supplies and, in the 50 days of confinement that followed, they were allowed to go out for a couple of hours a day to exercise, or to go to work if given the appropriate permission.

In a matter of months, a new terminology became part of the daily vocabulary of all Spaniards. Wording such as state of alarm, curfew, lockdown, perimeter closures or capacity restrictions are increasingly used, in Spain and abroad [3–7]. At the same time, new habits appeared—use of disinfectant gel and masks, increased interpersonal distancing, care of the immune system, attention to physical exercise and diet and so on. All this increased the general fear of the population, with a latent “fear” in the lives of many Spaniards—the fear of getting sick, of infecting, of the unknown.

This fear has caused changes in the use of public space worldwide [8–12] and leisure activities for Spaniards, especially with regard to their relationship with the restaurant

sector, in which Spain holds a unique position at a European and global level [13]. As discussed later, in Section 3.1, Spaniards have changed their habits related to spending time out of the house as well as the frequency with which they go to bars and restaurants, demonstrating a clear preference for open spaces over closed spaces.

The restaurant sector has not remained oblivious to these changes and has tried to adapt the offer to the new reality. This has caused a change in the appearance of some cities, where more bars and restaurants have opened new terraces which occupy pavements and public parking spaces [14,15]. Added to the new social and psychological reality is the impact of not being able to move freely between regions as a consequence of mobility restrictions. This has meant local restaurants of big cities have gained new local clients, helping to compensate for the additional costs of upgrading outside spaces. Such is the case of Madrid, the capital of Spain. In this study, we focus on the specific case of the city of Madrid because, in addition to presenting particular characteristics, as discussed in Section 2.3 of the paper, the local government has implemented a set of extraordinary fiscal measures to support the restaurant sector, so hard hit during the pandemic. The increase in terraces in public spaces raises the question of whether this policy initiative, introduced to help restaurants recuperate costs and losses caused by the total closure during confinement and subsequent capacity restrictions, is a short-term measure or here to stay, and whether it can serve as an example of economic recovery to other cities, in Spain and abroad.

From a methodological perspective, this paper exploits, using a statistical approach, both data available in official sources and our own collected data. On the one hand, official sources include data attained using traditional tools (such as probabilistic samples and administrative records) as well as data coming from pilot studies on mobility based on mobile phone positioning. On the other hand, we also rely on data collected using non-probabilistic surveys and on the use of ecological correlations. In particular, the research presented in this article is supported by data from official bodies such as the National Institute of Statistics (INE), the Statistical Office of the European Union (Eurostat) or the open data portal of the Madrid City Council in addition to other bibliographic sources, duly cited and referenced, as well as data about the impact of the COVID-19 pandemic collected using snowball sampling through web questionnaires. With all this, the uniqueness of the city of Madrid is revealed, whose special characteristics have led, together with government support and fostered by the human consequences of the pandemic (fear of accessing closed spaces without ventilation, limitations in movement, teleworking and resentful social relations), to an important increase, in number and extension, of the terraces in its streets.

The rest of the paper is structured as follows: Section 2, composed of four subsections, offers relevant background, contextualizing several aspects related to the topic of the paper. Section 2.1 presents a brief revision of the literature, emphasizing the relevance of the research, as this paper focuses on a topic not addressed to date. Section 2.2 provides a description of the restaurant sector in Spain, including its position in Europe. In Section 2.3 we take a detailed look at the singularities of the municipality (and Community) of Madrid, showing the reader the idiosyncrasy of the Madrilenian citizens, forced to modify their social habits in the face of such long perimeter confinement. Section 2.4 points out specific aspects of COVID-19 related to Madrid. Section 3 presents results and discusses them. Section 3.1, based on our own data collected through two specific web surveys, studies the impact that the COVID-19 pandemic has had on the social behaviour of Spaniards, and Section 3.2 discusses and shows the effects of the pandemic on the capital of Spain, focusing on the analysis of the supply of seating and terraces in the restaurant sector, answering several interesting questions. A short conclusion ends the paper.

2. Background

2.1. A Brief Revision of the Literature

Since the coronavirus pandemic began, numerous studies have been published on the impact of COVID-19, covering a multitude of topics. Research has mainly focused on issues related to health and, to a lesser extent, the economy, but issues related to cities

have also been addressed. We can find research that studies the impact of COVID-19 on urban air quality [16–19] or on the effects of human concentration, the consequences for public transportation or on new mobility patterns [20,21]. Noteworthy as well is the work on the need to transition to a more sustainable and resilient type of urbanization, which includes a change in the architecture of homes, creating more green spaces, with spaces that allow teleworking and the means for virtual education [22–24], or on the convenience of decongestion in urban habitats towards the surrounding rural spaces [25].

A minor line of literature has focused on the effects of COVID on restoration and on the economic impacts suffered by the sector [26]. Along these lines, a significant number of studies highlight the need for investment to improve the security of establishments (such as the use of contactless menu boards, payment systems, routine sanitization of tables, or screening of diners) and to have public support to do that [27–29]. Finally, we draw attention to the work of Alonso-Montolio [30] on the terraces of Barcelona but approached from the perspective of their energy consumption.

However, no research studies the change in the appearance of a city motivated by the boom of terraces, which in the case of Madrid is fostered by several factors, including citizens' fear of closed spaces, government support for them, with exemption from fees and the non-closure of establishments (when in other cities they were closed), and the high demand from their citizens, increased by mobility restrictions. These and other factors will be discussed throughout the article.

2.2. The Restaurant Sector in Spain

The restaurant sector is one of the drivers of Spain, both in an economic and social dimension. The hospitality industry, particularly the restaurant subsector, is one of the main references of the Spanish lifestyle [13]. In fact, this is one of the characteristics that differentiate Spain from other parts of the world and helps to put it in the top three tourist destinations in the world by number of visitors and revenue. According to the World Tourism Organisation (WTO) [31], Spain is the second most visited country in the world after France, and the second in tourist expenditures after the United States. Among the reasons given by foreign visitors are the excellent beaches, good weather and a wide range of gastronomy and nightlife [32–34].

However, according to Cabiedes and Miret-Pastor [35], it is difficult to offer exact figures of what the restaurant sector represents in Spain since, due to the diversity of the offer and a lack of regular and consistent data, the sources of information do not give a clear picture. Nevertheless, it is clearly a fundamental component of the Spanish tourism sector, the latter being of major economic importance. The tourism sector generates more than 2.6 million jobs, represents 12.8% of total employment and contributes 12.3% to the GDP. The contribution to the national GDP of the restaurant sector in particular is estimated to be 4.7%, employing 1,715,400 people in 2019, and representing 8.7% of the total employed in Spain [36] and more than 65% of the total employed in tourism.

Despite the fact that, every year, bars have been slowly decreasing in number, Spain still has approximately 280,000 establishments including bars, restaurants, cafes and community and catering companies [37], serving a population of just under 50 million people. In fact, since 2014 and coinciding with the end of the financial crisis which began in 2008, restaurants have shown a certain growth trend. In absolute terms, Spain is one of the countries in the European Union with the most bars and restaurants. The latest information collected by the Statistical Office of the European Union (Eurostat) in 2018 shows that Spain has 260,306 establishments, surpassed only by Italy, with 283,517 [38]. Considering the population of each country, Spain has a ratio of 557.90 establishments per 100,000 inhabitants, exceeded only by Portugal, Greece and Cyprus (see Figure 1), all of them Mediterranean countries.

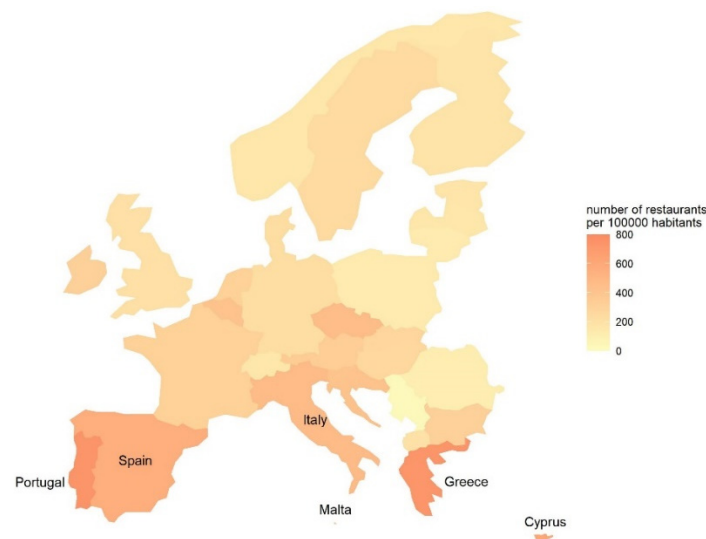


Figure 1. Number of restaurants in Europe per 100,000 habitants. Compiled by the authors using data from Eurostat [38]. The countries with the highest ratios are (i) Portugal (736.40), (ii) Greece (735.32), (iii) Cyprus (601.69), (iv) Spain (557.90), (v) Malta (470.46) and (vi) Italy (468.75).

Looking in more detail at the situation in Spain, Figure 2 shows that the region with the highest ratio of establishments per 100,000 inhabitants is the Illes Balears with 668.09, followed by the Islas Canarias with 627.96, Asturias 612.49 and Castilla y León with 603.99, all of them exceeding the ratio in Cyprus. The Community of Madrid ranks third from the bottom, with a ratio of 377.45 [39]. This fact, that of a lower relative offer in the Community of Madrid, together with the mobility restrictions and the significant demand characteristic of the people of Madrid for leisure and restaurants, is one of the factors that explain the boom in the increase in the number of restaurant terraces in Madrid.

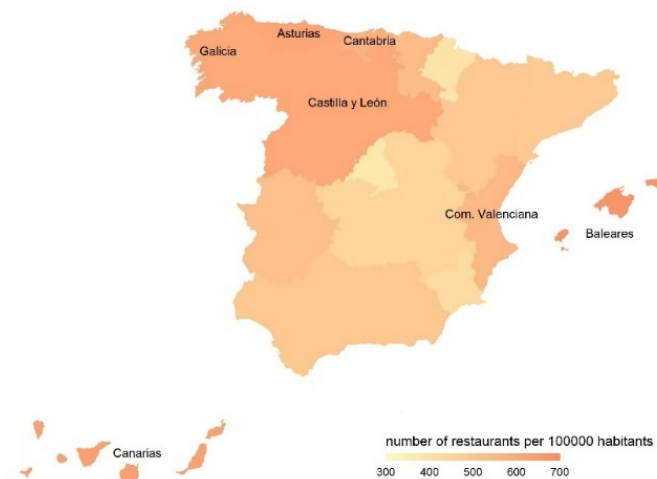


Figure 2. Number of restaurants in Spain per 100,000 habitants. Compiled by the authors using data from INE [39]. The regions with the highest ratio are Illes Balears (668.09), Islas Canarias (627.96), Asturias (613.49), Galicia (604.59), Castilla y León (603.99), Cantabria (568.87) and Comunitat Valenciana (545.42).

The COVID-19 outbreak has led to an unprecedented crisis in Spain [40] and particularly in the restaurant sector [41]. All the restrictions imposed to deal with COVID-19 have severely affected the economic activity of this sector, modifying the consumption behaviour of the population, both in bars and restaurants. Our study, focused on Madrid with its multiple singularities, looks at the response from this sector and its effects on the urban landscape.

2.3. Singularities of Madrid

Although the response of the restaurant sector (mainly a significant increase in the number and size of outside seating areas) has spread to a greater or lesser extent throughout the Spanish urban geography [42], we focus our analysis on Madrid, the capital city of Spain. This is because Madrid has specific characteristics different from other Spanish towns and cities. These characteristics, together with a different way of managing the health crisis by its authorities [43], means that bars and restaurants had more confidence in taking on the additional but necessary investment costs called for [44]. The economic and social dimension of the issue transcended to the political dimension, with the debate on the right to enjoy leisure time in bars or restaurants becoming the focus of the electoral campaign in the last Assembly of Madrid regional elections, held in May 2021.

What is it that makes Madrid so different? There are several characteristics, such as its population size, its relative wealth or its geolocation and demographic composition, which explain why the leisure behaviour of Madrid residents presents different characteristics from that of Spaniards as a whole.

Madrid has a high population. The state capital has more than 3.3 million inhabitants, a figure that exceeds 5 million if its metropolitan area is added, and almost 6.8 million if the region as a whole is taken into account [45]. This fact, together with its relatively small surface area, means it has the highest population density of all the Spanish regions, at 844.53 inhabitants/km², followed by the Basque Country, with a much lower density of 306.95 inhabitants/km² [45,46].

From an economic point of view, Madrid is also the richest region in Spain. Its gross domestic product per inhabitant is the highest in Spain at almost 36,000€, followed by the Basque Country with just over 34,000€, Navarra with 32,000€ and Cataluña with 31,000€ [47]. This higher population concentration and its greater relative wealth constitute, without a doubt, vectors that help to explain the greater mobility for leisure purposes of the people of Madrid—a mobility that is favoured by its geographical location, its transport infrastructures and its demographic composition.

Madrid is located in the geographical centre of the Iberian Peninsula, an ideal situation that favours travel. From Madrid it is possible to travel, with relative ease, to any part of the Spanish territory, not only because of the shorter distances but also because of an outstanding network of infrastructure. Madrid's airport has the highest number of flights and passengers in the whole country, and its air traffic (which accounts for 27% of that of the whole country) surpasses that of the Islas Canarias and the Islas Baleares [48]. However, its greatest strength is in its land transport network. As stated on the official website of the Community, "Madrid is the epicentre of the national road and rail network, being the best connected node in the country's transport network." From the capital it is possible to travel by high-speed train to tourist spots such as Toledo, Ciudad Real or Cuenca in less than an hour or to Valencia in just over an hour and a half. By road, six motorways (from A1 to A6) connect Madrid with other large Spanish cities (Bilbao, Barcelona, Valencia, Sevilla, Badajoz and La Coruña, respectively). Residents of Madrid can travel, through these routes, to any part of the country, in just a few hours.

In addition to the above data, we should consider the high percentage of "Madrileños not born in Madrid". The INE indicates, based on the continuous register, updated as of 1 January 2021 [45], that those born in other autonomous regions and registered as living in Madrid make up 25.4% of the total. Figure 3 shows how a significant percentage of the population are born in neighbouring autonomous communities such as Castilla y León and Castilla-La Mancha, followed by Andalucía and Extremadura. The excellent transport network of Madrid makes it easier for many residents in Madrid to travel regularly to their places of origin to visit their relatives, or to second homes. In fact, a high percentage of Madrid residents have a second home outside the autonomous community of Madrid. For example, in 2019 alone, more than 30,000 Madrid residents bought a home outside their autonomous community [49]. Madrid is the autonomous community that generates the most second home buyers, their purchase destinations of choice being, from largest to

smallest number, Toledo, Alicante, Malaga, Barcelona, Valencia, Almería, Cádiz, Murcia, Castellón, Guadalajara, Ávila, Sevilla, Las Palmas, Granada, Ciudad Real, Coruña, Girona, Cuenca, Zaragoza and others.

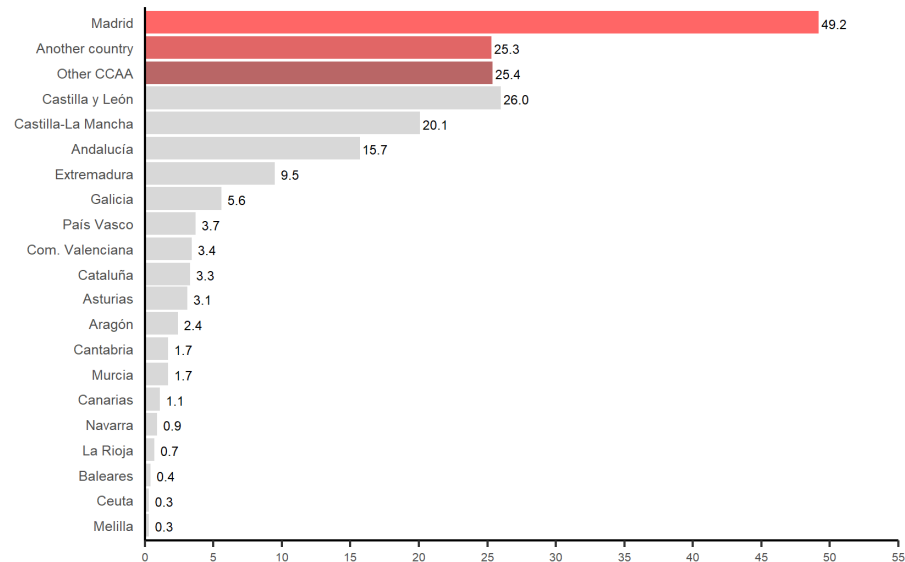


Figure 3. Distribution by autonomous communities of the percentage of allochthonous residents—those registered as living in Madrid born in other autonomous communities. Compiled by the authors based on data from the Permanent Municipal Register of the National Institute of Statistics, updated as of 1 January 2021 [50].

All of the above factors mean that the level of travel of residents of Madrid is higher than that of other Spanish citizens (see Figure 4). The higher relative income of Madrid residents is undoubtedly another factor. For Madrid residents, it is quite common to travel in and around Madrid for gastronomic leisure, to neighbouring provinces such as Guadalajara, Cuenca, Toledo, Ávila, Segovia, even Burgos and, of course, to different areas and municipalities within the Community of Madrid (Alcalá de Henares, Aranjuez, Chinchón, El Escorial, La Sierra de Guadarrama, Patones, etc.). The newspaper headlines after confinement illustrate this fact, for example, “Segovia recovers its tourists from Madrid” [51], “Welcome Madrid, Guadalajara A-2 pasos (two steps away)” (this was a campaign launched by the Government of Castilla-La Mancha in 2008 and relaunched after curfew was lifted), “The people of Madrid are still very much present in Aranda del Duero” [52].

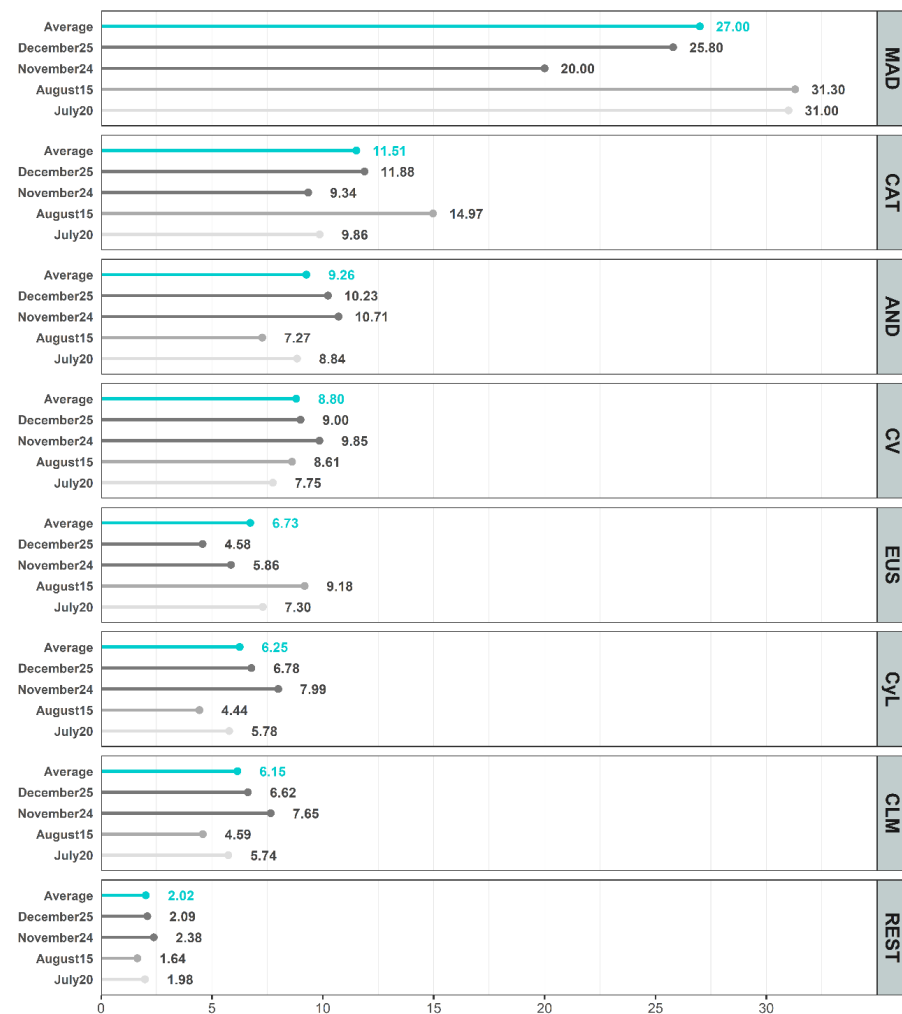


Figure 4. Mobility of residents to other autonomous communities. Compiled by the authors using data from the INE corresponding to the pilot study, “Mobility studies using mobile telephones” [53]. The percentage refers to the number of travellers from the reference community to other communities with respect to the total number of travellers per day. The acronyms refer to the following autonomous communities: MAD (Madrid), CAT (Cataluña), AND (Andalucía), CV (Comunitat Valenciana), EUS (País Vasco), CyL (Castilla y León) and CLM (Castilla La Mancha). REST (the rest of the autonomous communities) aggregates the average of the remaining 12 Spanish regions. The dates chosen correspond to public holidays.

Madrid residents are the most avid travellers in Spain. In 2020 alone, they accounted for 17.2% of total journeys completed in Spain, followed by residents of Andalucía with 16.4% (being more than 1.5 million more inhabitants) and Cataluña with 16.1% (being 1 million more inhabitants). Madrid residents also lead the way in the number of overnight stays made during their trips, with 23.3% of the total, accounting for 22.5% of spending [54].

Moreover, taking into account the data provided by the INE through a pilot study, “Mobility studies using mobile telephones” [55], we were able to verify that on the four public holiday dates in 2019 on which the movement of travellers was monitored (20 July, 15 August, 24 November and 25 December), the autonomous community of Madrid was the one with the highest percentage of travellers outside its autonomous community. Figure 4 shows that, on each of the four days, the autonomous community of Madrid had the highest number of travellers, followed by Cataluña and Andalucía (both with greater populations than Madrid) and the Comunitat Valenciana, the Basque Country and both Castilla-La Mancha and Castilla y León. The remaining Spanish autonomous communities, combined, did not exceed the data for Madrid.

According to the data of the aforementioned study carried out by the INE, 25% of the trips made by Madrid residents were destined for the autonomous community of Castilla-La Mancha, while another 20% were to Castilla y León (see Figure 5), its border regions. However, the journeys were not only to the autonomous communities that surround the Madrid region. Trips to Andalucía and the Comunitat Valenciana were also significant, adding up to 13% and 8%, respectively. The remaining 34% travelled to other parts of the Spanish peninsula, helped by the particular geographical location of the capital, as mentioned above, and the extensive infrastructure and transport networks that Madrid has to offer.

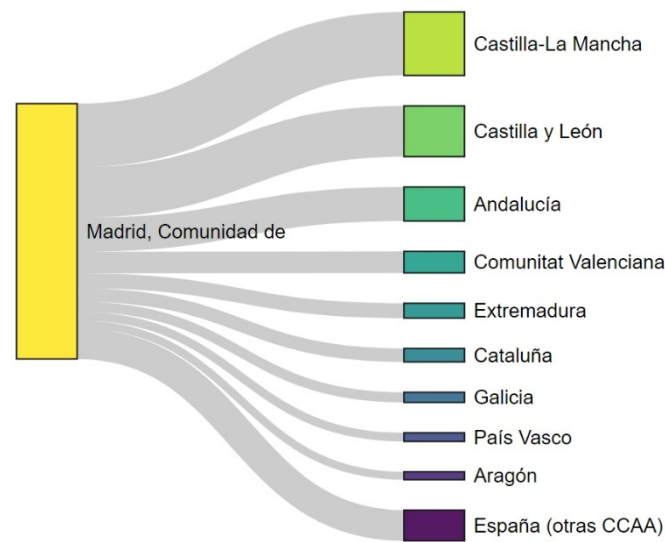


Figure 5. Mobility of Madrid residents on 24 November 2019 to other autonomous communities. Compiled by the authors using data from the INE [45] corresponding to the experimental study, “Mobility studies using mobile telephones”. The figure shows the number of Madrid residents who travelled to each autonomous community with respect to the total number of travellers from Madrid that day. From highest to lowest percentage: Castilla-La Mancha (24.9%), Castilla y León (19.9%), Andalucía (13.3%), Comunitat Valenciana (8.4%), Extremadura (5.8%), Cataluña (5.2%), Galicia (4.0%), País Vasco (3.4%), Aragón (3.1%), Islas Canarias (2.7%), Murcia (1.9%), Cantabria (1.7%), Navarra (1.2%), Islas Baleares (1.2%), La Rioja (1.1%), Melilla (0.1%) and Ceuta (0.1%).

The reasons mentioned above show Madrid as an autonomous community with a population that embraces travel—Madrid residents are highly mobile due to their geographical location, their income, their easy access to the transport network and their demographic and social profile. The region is made up of many “Madrileños from outside Madrid” who tend to make frequent visits to their destinations of origin and to their second homes, located throughout the Spanish peninsula. In addition, Madrid is a city that “lives in the street”, with very long business hours, including heightened activity on Sundays and bank holidays with its liberalised trading hours [55].

2.4. COVID-19. Restrictions and Changes in Habits

The mobility patterns outlined in the previous subsection have been drastically modified due to the long period of perimeter closure that almost all the Spanish regions (excluding Madrid) decreed for six months. On 25 October 2020, the Government of Spain approved, by Royal Decree, the third state of alarm in less than a year, which ended on 9 May 2021 [56]. This extraordinary measure gave the different Spanish regional authorities the power to close their territory to visitors from other regions. This fact meant that the Spanish population was only able to move within its territorial limits, which, for the people of Madrid, imposed a significant limitation. Added to its high population density and fondness for travel is the particularity of being a single-province autonomous community,

a fact that increased the feeling of confinement perceived by residents. This feeling was very different from that experienced by citizens of other autonomous communities, whose inhabitants had the possibility of travelling between provinces.

The restriction of mobility beyond its borders meant the city of Madrid became the preferred focus of leisure for residents in the region. This, coupled with the limitations imposed for health reasons on using indoor space in bars and restaurants and the fear felt by a significant part of the population of sitting inside restaurants, led to a change in habits and in the appearance of the city's streets.

The limitation placed on the capacity in restaurants and hotels, together with the need of the resident population in Madrid for leisure spaces, has motivated the local government to lend support to this sector by offering financial assistance and by allowing bars and restaurants to add or extend outdoor terraces in order to meet public demand. This authorization, without additional cost to the bars and restaurants, has facilitated the occupation of public spaces. A subsidy of 100% of the usual charge has meant establishments could open terraces, at no extra cost, for at least two years, 2020 and 2021 [57,58].

All these factors led to Madrid expanding recreational spaces in order to meet an important demand for interior tourism in times of mobility restrictions.

3. Results and Discussion

3.1. The COVID-19 Pandemic and the Impact on the (Leisure) Behaviour of Spaniards

Between 28 April and 14 May 2020, the Research Group on Electoral Processes and Public Opinion of the University of Valencia (GIPEyOP), of which the authors of this research are members, launched (in full confinement) a survey with a view to understanding the perception and assessment by Spaniards of the situation they were experiencing [59]. The survey collected 8387 valid responses through a snowball sample design, initiated from a file of GIPEyOP collaborators. This type of sample design, which is not probabilistic but chained, allows quality information to be obtained with the right appropriate processing [60,61]. The survey was sent by email or instant messaging to collaborators, and they, in turn, forwarded it to their contacts. The individual responses obtained were weighted using post-stratification/calibration techniques to correct biases in the collected sample [62].

Given that the bulk of responses were obtained during the first few days after the survey was launched (92% before 2 May, the first day when people were allowed, by time slots, to go outside for physical exercise, and 96.8% during the first week), we can affirm that the responses obtained were drawn from a situation of home confinement. Hence, from now on we refer to this survey as the Lockdown Survey (LS).

The survey was answered mostly by women, 54.3%, a significant difference with respect to what usually happens in the surveys carried out by GIPEyOP, mostly answered by men [63]; without doubt the confinement situation help this fact. The age distribution showed the mean age to be 50.7 years with a standard deviation of 14.1, as can be calculated from the figures shown in Table 1. The bulk of respondents were aged between 45 and 64 years, while the age group with the lowest relative representation was that of those over 64 years of age, since despite the fact that their percentage in the sample is higher than that of the younger group (17.4% compared to 10.8%) the former group represents a higher percentage of the population.

Table 1. Age distributions (in percentages) in the GIPEyOP surveys analysed.

	18–30	31–44	45–64	Over 64
Lockdown Survey	10.8	21.5	50.2	17.4
Post-Lockdown Survey	11.8	22.8	46.2	19.2

Source: compiled by the authors using results from the Lockdown Survey [59] and Post-Lockdown Survey [64].

One question, which was answered by 88.1% of the sample, asked respondents whether they had been out of the house during the weeks of confinement and how they

felt. The survey revealed that, in general, the population went out “only when necessary” and “without fear”. However, it is significant that 37.4% of respondents did say they were afraid to go out. In fact, 9.4% stated that they did not go out, and 28% responded that they went out, but with fear (see Figure 6).

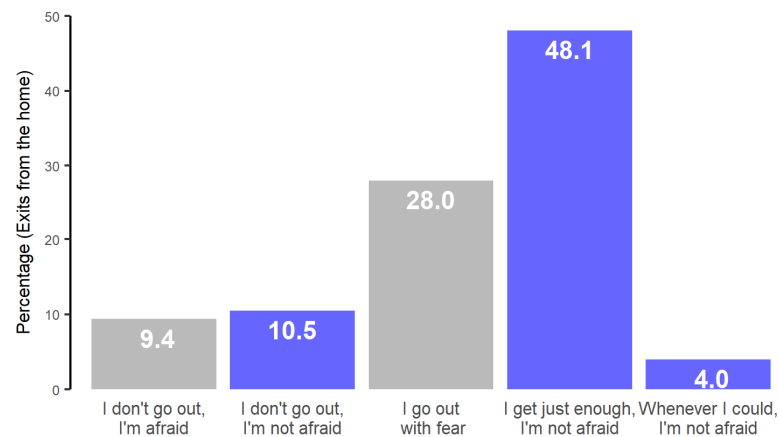


Figure 6. Fear and leaving the house. Lockdown Survey [59]. The options to choose when answering the question were the following: (i) I have not gone out during the entire period of confinement and I am afraid to do so; (ii) I have not gone out during the entire period of confinement, but I am not afraid; (iii) I go out to do errands (walking the dog, shopping, work, care . . .) and I do so with fear; (iv) I have gone out only when necessary, to go shopping and/or to go to work and I am not afraid; and (v) I have gone out whenever I could and I am not afraid.

The fear of leaving home and being infected, not only of Spaniards [65–67], together with other factors, such as changes in social habits, not being able to visit friends and family, as well as the uncertainty about the national economy future or the family economy, can explain why, in line with other studies [68,69], 49% of those surveyed claimed they had suffered changes in their sleep habits during the pandemic.

A few months after carrying out the Lockdown Survey (LS), the GIPEyOP researchers conducted a second survey [64], using the same channel (social media) and procedure (snowball or chain sampling). The investigation of this second survey was framed in the period between 23 September and 14 October 2020, coinciding with the so-called new normal life and with the beginning of the second wave of the pandemic in Spain. Hereafter, we refer to this second study as the Post-Lockdown Survey (PLS). This second survey, which collected 1955 valid responses, was carried out at the end of the summer, a period during which Spain experienced a certain relaxation of most of the restrictive measures in place over the previous months.

In this case male participation, of 56.1%, was somewhat higher, although the age distribution was quite similar to that obtained in the LS survey. As can be derived from Table 1, the mean age of the respondents was 50.6 with a standard deviation of 15.7 years of age. The age variable presented a greater dispersion in the PLS.

In this second survey, which asked questions related to the attitude with which people faced the new situation and the possible changes (or not) experienced as a result of COVID-19, significant alterations were seen in various dimensions, including labour and social dimensions.

With regards to the work environment, the survey highlighted significant effects in terms of job performance and the way of working, where teleworking had a strong presence. In the PLS, 33.6% of the worker respondents answered that they performed less at work, 47.9% the same and only 18.4% more. Of those surveyed, 7.9% were teleworking, and 13.5% alternated between the two systems (face-to-face work and telework). Of those remotely working, 30.9% would choose to continue to do so, surely because the fear established in the society since the beginning of the pandemic [8,65,70]. However, 40.7% preferred to alternate both work modalities, on the one hand, due to a social need to interact or

communicate in a more personal way, and on the other hand, to cut down on commuting to work, saving on time and money and reducing also the risk of infection [71–75].

Although in the first survey, during confinement, 37.4% said they were afraid, in this second survey, after confinement and restrictions had been lifted, this percentage increased to just over 50%. A total of 2.8% said they had hardly left the house out of fear, 13.4% said they went out to do basic errands but were afraid and 34.6% said they left the house as normal but were afraid (see Table 2).

Table 2. Percentage distribution of the question about leaving the house.

Talking about Leaving the House after Confinement	Percentage
I have hardly gone out, but I am not afraid	5.3
I have hardly left the house out of fear	2.8
I go out alone to do basic errands, although I am afraid	13.4
I go out to do basic errands, and I am not afraid	25.0
I go out as normal, and I am not afraid	18.8
I go out as normal, but I am afraid	34.6

Source: compiled by the authors using results from the Post-Lockdown Survey [64].

Tables 3 and 4 show that there were modifications, both in relationships and in habits, with regards to leaving the house. In the former, only 13.1% affirmed that their behaviour was the same in terms of their relationships. In the latter, that of habits, only 12.6% affirmed that they did not restrict themselves when going out, while more than half, 55.3%, stated that they tended to avoid closed spaces (bar and restaurant interiors).

Table 3. Distribution (in percentages) of the change in social relationships.

Change in Relationships as a Result of COVID-19	Percentage
No, same as before	13.1
Yes, I only interact with people I live with	5.5
Yes, I only interact with a close circle of contacts	37.6
Yes, I have reduced the number of people I interact with	43.8

Source: compiled by the authors using results from the Post-Lockdown Survey [64].

Table 4. Distribution (in percentages) of the change in leaving the house.

Change in Relationships as a Result of COVID-19	Percentage
Modification of habits in leaving the house	22.1
I choose closed spaces as long as there is good ventilation	10.0
I usually avoid closed spaces	55.3
I do not restrict myself in my choice of spaces	12.6

Source: compiled by the authors using results from the Post-Lockdown Survey [64].

The data in Tables 3 and 4 clearly show a reduction in the number of times people leave the house and a noticeable change in where they choose to go. The effects of these changes on restaurants and other eating places are/have been devastating. In fact, many have had to adapt their services to subsist [76,77]. Before the state of alarm, practically everyone met in a bar or restaurant; 94.5% did so at least once a week, with only 5.5% of those surveyed saying they never did (see Figure 7). Since confinement and the state of alarm, the size of this latter group has soared, with more than a third of the people surveyed, 36.6%, saying they avoided going to bars and restaurants after the state of alarm. In less than six months, the average weekly number of outings to bars and restaurants of the Spanish population more than halved (from 2.69 to 1.32 times).

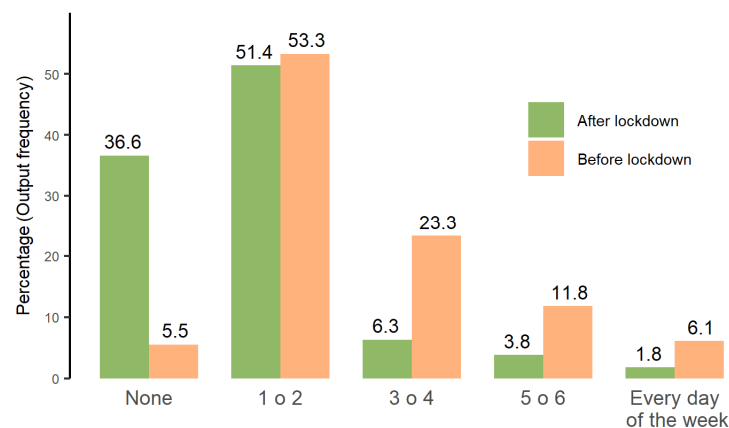


Figure 7. Distribution (in percentages) of frequency of going to bars and restaurants. Post-Lockdown Survey [64]. The question was how often did you/do you usually go to a bar or restaurant (before and after the state of alarm). The options to choose from to answer the question were the following: (i) Not on any day; (ii) 1 or 2 days a week; (iii) 3 or 4 days a week; (iv) 5 or 6 days a week; and (v) Every day of the week.

We should keep in mind that this manifestation of fear, both of leaving the house and of going to bars and restaurants, especially if inside, is a psychological response to a home confinement that lasted more than 3 months, and a high degree of uncertainty about the way the virus was passed on and its effects [9,65,70]. The consequence of this for the restaurant sector extended beyond the lockdown and it became clear that, in order to survive, the sector would need to reinvent itself.

3.2. Discussion: The Boom of Restaurant Terraces

In light of the above, the following questions arise: How has the service offered by the restaurant sector in Madrid been affected? Has the distribution of terraces in the city changed by districts compared to a year ago? To answer these questions, we used the data available in the Madrid City Council Open Data Portal [78]. Specifically, we compared the data for April 2018 (before COVID-19) with the latest published data (April 2021).

Although we worked with official data, we can see from our own observations that, in reality, there are more terraces than those recorded by these sources, and more than reported by the media. For example, the digital newspaper *El Mundo*, on Thursday, 20 May 2021 [79], states that that the City Council had authorized, on an exceptional and provisional basis, about 3000 terraces called COVID terraces, 582 of them occupying parking spaces. In fact, according to other written media, the number of public parking spaces occupied by terraces is 1502 (1328 of them in green zones, parking areas set aside for residents), the Chamberí district being the most affected, with 507 (450 in the green zone and 57 in the blue zone), followed by the Salamanca district (397 in the green zone) [80]. According to local government data, in April 2018 there were 4879 licensed terraces, while in April 2021 there were 6275, that is, 1396 more terraces, of which 517 occupied public parking spaces.

Even working with underestimated official figures, we observed that over the three-year period of 2018–2021, there was an increase in seating capacity in all districts of the capital. As Figure 8 (right side) shows, all districts saw an increase in the seating capacity granted due to the exceptional measures in place because of COVID-19.

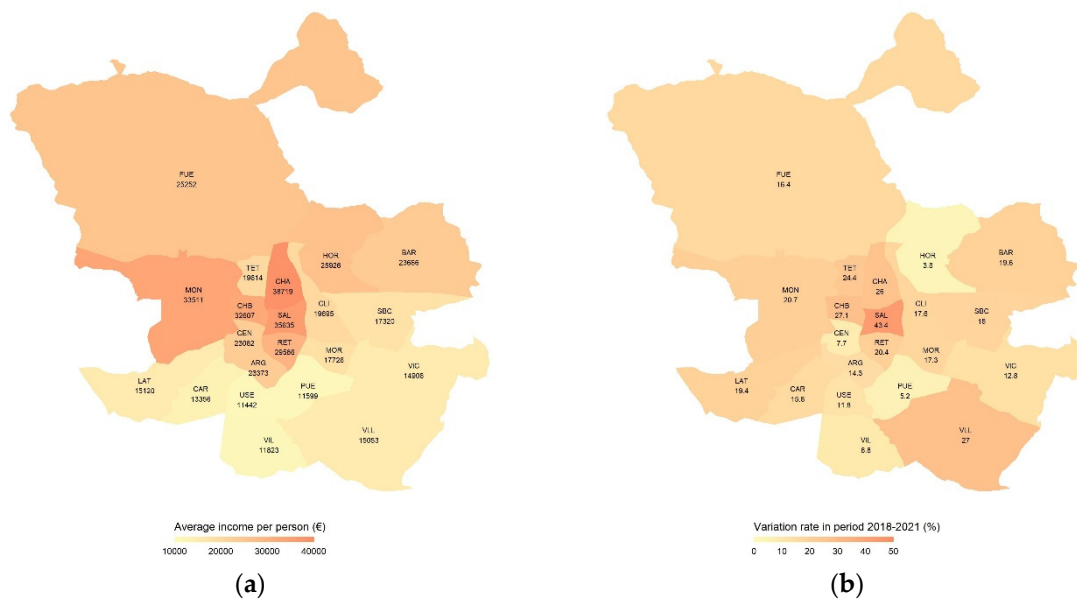


Figure 8. (a) Per capita income (in euros) of Madrid residents; (b) variation rate (in percentage) of the terrace seating capacity (period 2018–2021). Compiled by the authors using data from the INE [81] and the Madrid City Council’s open data portal [78].

The licensing of new terraces, or the expansion of the seating available in existing ones, is not in line with the population of the districts but rather with their per capita income, as reflected by comparing the two representations in Figure 8. It is in the richest districts of the city (Chamartín, Salamanca, Chamberí and Moncloa) where there has been a greater increase in terrace seating capacity. In fact, a strong correlation (0.55) is observed between per capita income and the increase in terrace seating capacity, which is consistent with the hypothesis that the higher the income, the greater the mobility. If the correlation between the increase in seating and the population is calculated, the value is negative (-0.21), that is, the fact that a district has a higher population has not led to an increase in terrace seating capacity. Mobility restrictions, therefore, have favoured a greater increase in the supply of terraces and seating in those places where residents have been most affected in terms of habits by these measures.

Having verified this change, the next question would be whether this new trend is here to stay. The Madrid City Council announced that on 1 January 2022 it will withdraw the licences for terraces with permission to occupy public parking spaces, maintaining, until further notice, those terraces with permission to occupy pavements [82]. Madrid City Council considers that this period gives sufficient time for establishments to recover investments made, such as in the installation of decking or the total or partial enclosure of a terrace, as well as in the acquisition of outdoor heating systems. Interestingly, the Mahou San Miguel beer company has tripled its investment in tables and chairs, allocating 2.6 million euros to offering versatile furniture to the more than 10,000 bars and restaurants in the capital, as part of its Global Support Plan [83], which includes not only furniture but also one-off contributions of beer and water to establishments.

The increase in seating on the streets of Madrid may be reduced in the coming months. How will this reduction affect each of the districts? According to available data (April 2021), the local government has authorised the placement of 8826 seats in parking areas. The distribution, by district, of this variable is graphically represented in Figure 9 (left panel). The districts of Chamberí, Salamanca, Arganzuela and Chamartín will be the most affected by the entry into force of the measure imposed by the authorities, since they have, respectively, 2580, 1718, 1083 and 1039 seats placed in public parking areas. However, also other districts, such as Usera, will be greatly affected by this measure. In fact, at the

time of writing this paper, 9.4% of the authorised seats in this Madrid area were located in public parking spaces.

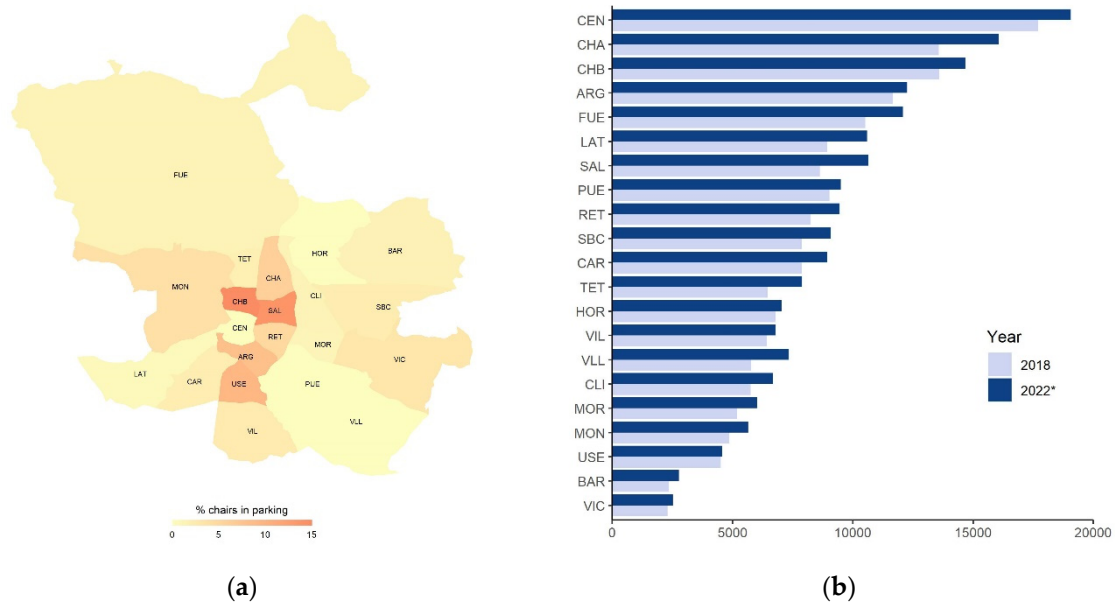


Figure 9. (a) Proportion of seats (on restaurant terraces) authorised to occupy public parking areas (April 2021) with respect to the total number of seats authorised in each district; (b) number of seats (on restaurant terraces) in April 2018 (blue light) and estimated for 2022, taking into account the data available in April 2021, and the seats that occupy parking spaces (dark blue). Compiled by the authors using data from Madrid City Council’s open data portal [78]. “*” stands for estimated (not observed) data.

However, as previously stated, the number of terraces in the capital of Spain has increased since 2018, and, although permission for seats that occupy public parking areas might eventually be withdrawn, the increase since 2018 would still be remarkable in every district of Madrid (see Figure 9, right panel).

In fact, the report in the newspaper *El Mundo* of 20 May 2021 implies that this measure (rescinding the licences of the terraces that occupy public parking) may not be applied, at least not on the date initially stipulated (January 2022). Several political parties present on the city council, such as *Ciudadanos* or *Más Madrid*, advocate that the terraces remain. The first point they make is that “there are things that are here to stay, like in New York” (Begoña Villacís). In their second point, they indicate that “the public space occupied by cars is excessive (. . .) although it is true that space taken from car parking should be used for other activities, not just for terraces” (Rita Maestre).

4. Conclusions

The restaurant sector is one of the most important drivers of the Spanish economy and is also one of the sectors that has suffered the most from the consequences of COVID-19. The pandemic has caused a series of changes in the habits and behaviour of citizens, ranging from the way people socialise to the way they work.

The fear of infection and spread of the virus has increased the demand for outdoor leisure spaces. The increase in teleworking means people are keen to leave the house at the end of the working day in order to establish/maintain social contact. Mobility restrictions have drastically reduced the number of long-distance leisure trips. All of the above has increased the demand for recreation areas nearer to home. The combination of these factors has acted as a catalyst, prompting the rapid adaptation of the restaurant sector in a city such as Madrid, whose characteristics have helped create ideal conditions.

Following an analysis of the above factors and studying the idiosyncrasies of Madrid, we use data to show what the impact has been on the city’s restaurant sector and on the

urban landscape, paying particular attention to its differential spatial distribution. The positive response of Madrid and its citizens in supporting the demand for space outside bars and restaurants, with the issue gaining a high profile in the public debate of the most recent electoral campaign, leads us to think that these changes will leave their mark, with a significant proportion of the terraces (and/or increased seating capacity) being maintained as part of the city landscape. Finally, the political and fiscal measures implemented in the Spanish capital have made it possible to revive the hospitality sector at a very difficult time. We believe that the case of Madrid could serve as an example for other Spanish cities and cities abroad.

Author Contributions: Conceptualization, J.M.P.; methodology, V.P., C.A. and J.M.P.; software, V.P. and C.A.; validation, V.P., C.A. and J.M.P.; formal analysis, V.P. and C.A.; investigation, V.P., C.A. and J.M.P.; resources, C.A. and J.M.P.; data curation, V.P. and C.A.; writing—original draft preparation, V.P. and C.A.; writing—review and editing, V.P., C.A. and J.M.P.; visualization, C.A. and J.M.P.; supervision, J.M.P.; project administration, J.M.P.; funding acquisition, C.A. and J.M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by SUPERA COVID-19 FUND, grant number ROCOGIS, and The APC was funded by SUPERA COVID-19 FUND.

Informed Consent Statement: At the beginning of the questionnaire in each online surveys, participants were informed that the survey was anonymous, voluntary, and confidential, as established in the current regulations on Personal Data Protection and guarantee of digital rights. It was also indicated that the conclusions drawn from the survey would be only presented in aggregate form.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to thank Marie Hodkinson for translating the text of the paper into English and the support of the SUPERA COVID-19 FUND, a joint tender offered by CRUE (Conference of Rectors of Spanish Universities), CSIC (the Spanish National Research Council) and Banco Santander (Santander Bank), through the project ROCOGIS (The Faces of COVID-19. Gender and Socioeconomic Impacts). The authors also acknowledge the support of Generalitat Valenciana through project AICO/2021/257 (Conselleria d'Innovació, Universitats, Ciència i Societat Digital).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Boletín Oficial del Estado (BOE). Real Decreto 463/2020, of 14 March, Declaring a State of Alarm in Order to Manage the Health Crisis Situation Caused by COVID-19. 2020. Available online: <https://www.boe.es/eli/es/rd/2020/03/14/463> (accessed on 12 August 2021).
2. COVID-19 Government Response Tracker. Blavatnik School of Government & University of Oxford. Available online: <https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker> (accessed on 12 August 2021).
3. González-Rodríguez, A.; Labad, J. Mental health in times of COVID: Thoughts after the state of alarm. *Med. Clínica* **2020**, *155*, 392–394. [[CrossRef](#)]
4. Ertugrul, S.; Soylemez, E. Explosion in hearing aid demands after Covid-19 outbreak curfew. *Eur. Arch. Oto-Rhino-Laryngol.* **2020**, *278*, 843–844. [[CrossRef](#)]
5. Devi, S. Travel restrictions hampering COVID-19 response. *Lancet* **2020**, *395*, 1331–1332. [[CrossRef](#)]
6. Zajenkowski, M.; Jonason, P.K.; Leniarska, M.; Kozakiewicz, Z. Who complies with the restrictions to reduce the spread of COVID-19: Personality and perceptions of the COVID-19 situation. *Personal. Individ. Differ.* **2020**, *166*, 110199. [[CrossRef](#)] [[PubMed](#)]
7. Martínez-Riera, J.R.; Gras-Nieto, E. Home care and COVID-19. Before, in and after the state of alarm. *Enfermería Clínica* **2021**, *31*, S24–S28. [[CrossRef](#)]
8. Liu, Q.; Zhao, G.; Ji, B.; Liu, Y.; Zhang, J.; Mou, Q.; Shi, T. Analysis of the influence of the psychology changes of fear induced by the COVID-19 epidemic on the body. *World J. Acupunct.–Moxibustion* **2020**. [[CrossRef](#)]
9. Rahman, M.A.; Hoque, N.; Alif, S.M.; Salehin, M.; Shariful-Islam, S.M.; Banik, B.; Sharif, A.; Nazim, N.B.; Sultana, F.; Cross, W. Factors associated with psychological distress, fear and coping strategies during the COVID-19 pandemic in Australia. *Glob. Health* **2020**, *16*, 95. [[CrossRef](#)]
10. Honey-Rosés, J.; Anguelovski, I.; Chireh, V.K.; Daher, C.; Konijnendijk, C.; Litt, J.S. The impact of COVID-19 on public space: An early review of the emerging questions—design, perceptions and inequities. *Cities Health* **2020**. [[CrossRef](#)]
11. Jasiński, A. Public space or safe space—remarks during the COVID-19 pandemic. *Tech. Trans.* **2020**, *117*, e2020020. [[CrossRef](#)]

12. Wang, D.; Yao, Y.; Brett, A.S. The effects of crowdedness and safety measures on restaurant patronage choices and perceptions in the COVID-19 pandemic. *Int. J. Hosp. Manag.* **2021**, *95*, 102910. [[CrossRef](#)]
13. Gallego, E. El sector de la restauración en España. *Distrib. Y Consumo* **2018**, *28*, 26–30.
14. Florida, R.; Rodríguez-Pose, A.; Storper, M. Cities in a post-COVID world. *Urban Stud.* **2021**. [[CrossRef](#)]
15. Rodrigo-Comino, J.; Senciales-González, J.M. A regional geography approach to understanding the environmental changes as a consequence of the COVID-19 lockdown in highly populated Spanish cities. *Appl. Sci.* **2021**, *11*, 2912. [[CrossRef](#)]
16. Bedi, J.S.; Dhaka, P.; Vijay, D.; Aulakh, R.S.; Gill, J.P. Assessment of Air Quality Changes in the Four Metropolitan Cities of India during COVID-19 Pandemic Lockdown. *Aerosol Air Qual. Res.* **2020**, *20*, 2062–2070. [[CrossRef](#)]
17. Kerimray, A.; Baimatova, N.; Ibragimova, O.P.; Bukenov, B.; Pavel, P.; Karaca, F. Assessing air quality changes in large cities during COVID-19 lockdowns: The impacts of traffic-free urban conditions in Almaty, Kazakhstan. *Sci. Total Environ.* **2020**, *730*, 139179. [[CrossRef](#)]
18. Lanchipa, T.M.; Moreno-Salazar, K.; Luque-Zúñiga, B. Perspectiva del COVID-19 sobre la contaminación del aire. *Rev. Soc. Científica Parag.* **2020**, *25*, 155–182. [[CrossRef](#)]
19. Nie, D.; Shen, F.; Wang, J.; Ma, X.; Li, Z.; Ge, P.; Ou, Y.; Jiang, Y.; Chen, M.; Chen, M.; et al. Changes of air quality and its associated health and economic burden in 31 provincial capital cities in China during COVID-19 pandemic. *Atmos. Res.* **2021**, *249*, 105328. [[CrossRef](#)]
20. Cenecorta, A.X.I. La ciudad que quisiéramos después de COVID-19. *ACE Archit. City Environ.* **2020**, *15*. [[CrossRef](#)]
21. Otaola, P. Oportunidades de las ciudades tras la crisis del COVID. *Rev. Obras Públicas Órgano Prof. Los Ing. Caminos Canales Y Puertos* **2020**, *3621*, 56–61.
22. Salvador-Ferrín, M. Transformación de la arquitectura desde el COVID-19. *Rev. Científica Y Arbitr. Obs. Territ. Artes Y Arquít. FINIBUS* **2020**, *3*, 26–45.
23. Mattioli, L.; Schneider, M.C. Redefiniendo nuestro futuro: La transformación de nuestras ciudades frente a la crisis COVID-19. In *Cities and COVID-19: New Directions for Urban Research and Public Policies*, 1st ed.; Delgado-Ramos, G.C., López-García, D., Eds.; Plataforma de Conocimiento para la Transformación Urbana: Mexico City, Mexico, 2020; pp. 340–345.
24. Luna-Nemecio, J.M.; Tobón, S. Urbanización sustentable y resiliente ante el Covid-19: Nuevos horizontes para la investigación de las ciudades. *Univ. Soc.* **2021**, *13*, 110–118.
25. Ferrás-Sexto, C. Cidades dispersas e aldeias virtuais no pós-pandemia da Covid-19. *Finisterra* **2021**, *55*, 243–248. [[CrossRef](#)]
26. Panzone, L.A.; Larcom, S.; She, P.-W. Estimating the impact of the first COVID-19 lockdown on UK food retailers and the restaurant sector. *Glob. Food Secur.* **2021**, *28*, 100495. [[CrossRef](#)]
27. De la Cruz-May, S.; May-Guillermo, E.G. Innovation practices implemented by MSMEs in the restaurant sector in the fase of COVID-19 in Tabasco, Mexico. *Nova Sci.* **2021**, *13*. [[CrossRef](#)]
28. Dube, K.; Nhamo, G.; Chikodzi, D. COVID-19 cripples global restaurant and hospitality industry. *Curr. Issues Tour.* **2021**, *24*, 1487–1490. [[CrossRef](#)]
29. Hobbs, J.E. The Covid-19 pandemic and meat supply chains. *Meat Sci.* **2021**, *181*, 108459. [[CrossRef](#)] [[PubMed](#)]
30. Alonso-Montolio, C.; Serra-Coch, G.; Isalgue, A.; Coch, H. The Energy Consumption of Terraces in the Barcelona Public Space: Heating the Street. *Sustainability* **2021**, *13*, 865. [[CrossRef](#)]
31. World Tourism Organisation. Impact of COVID-19 on Global Tourism Made Clear as UNWTO Counts the Cost of Standstill. 2020. Available online: <https://www.unwto.org/news/impact-of-covid-19-on-global-tourism-made-clear-as-unwto-counts-the-cost-of-standstill> (accessed on 12 August 2021).
32. Hosteltur. Las 10 Razones Que Explican el Éxito de España Como Destino Turístico. 2017. Available online: https://www.hosteltur.com/120178_10-razones-explican-exito-espana-como-destino-turistico.html (accessed on 12 August 2021).
33. Ostelea. Ocio Nocturno a Nivel Global: Un Fenómeno de Dinamización Económica. 2019. Available online: http://www.aept.org/archivos/documentos/informe_ostelea_turismo_y_ocio_nocturno.pdf (accessed on 12 August 2021).
34. Hosteltur. El 23% del Turismo Que Llega a España Viene Por el Ocio Nocturno. 2019. Available online: https://www.hosteltur.com/comunidad/nota/018719_el-23-del-turismo-que-llega-a-espana-viene-por-el-ocio-nocturno.html (accessed on 12 August 2021).
35. Cabiedes, L.; Miret-Pastor, L. Fuentes estadísticas para analizar el sector de la restauración en España. *Papers* **2019**, *104*, 129–145. [[CrossRef](#)]
36. Instituto Nacional de Estadística (INE). Economically Active Population Survey. Available online: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=ultiDatos&idp=1254735976595 (accessed on 12 August 2021).
37. Hostelería de España. Anuario de la Hostelería de España 2020. Available online: <https://apehl.org/media/editor/2020/12/15/anuario-2020.pdf> (accessed on 12 August 2021).
38. Eurostat. Annual Detailed Enterprise Statistics for Services. Available online: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=sbs_na_1a_se_r2&lang=en (accessed on 12 August 2021).
39. Instituto Nacional de Estadística (INE). Statistical Use of the Central Business Register. CBR. Available online: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736160707&menu=ultiDatos&idp=1254735576550 (accessed on 12 August 2021).
40. Cabrer-Borrás, B.; Rico, P. Impacto Económico del Sector Turístico en España. *Stud. Appl. Econ.* **2021**, *39*, 1–19. [[CrossRef](#)]

41. Moreno-Luna, L.; Robina-Ramírez, R.; Sánchez-Oro Sánchez, M.; Castro-Serrano, J. Tourism and Sustainability in Times of COVID-19: The Case of Spain. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1859. [CrossRef] [PubMed]
42. Uría Menéndez Lawyers. Spain's Response to COVID-19: Emergency Measures; Gradual Relaxation. *International Financial Law Review*. 2020. Available online: <https://www.iflr.com/article/b1lxmrrfr4gkfs/spains-response-to-covid19-emergency-measures-gradual-relaxation> (accessed on 12 August 2021).
43. Madrid City Council Website. El Ayuntamiento de Madrid Aprueba un Paquete de Nuevas Medidas Fiscales Con Una Rebaja de 107,5 Millones de Euros Para 2021. Available online: <https://www.madrid.es/portales/munimadrid/es/Inicio/Actualidad/Noticias/El-Ayuntamiento-de-Madrid-aprueba-un-paquete-de-nuevas-medidas-fiscales-con-una-rebaja-de-107-5-millones-de-euros-para-2021/?vgnnextfmt=default&vgnnextoid=0c26bc6d6c375710VgnVCM2000001f4a900aRCRD&vgnnextchannel=a12149fa40ec9410VgnVCM100000171f5a0aRCRD> (accessed on 27 August 2021).
44. Urrutia, C. Madrid Lidera la Recuperación del Empleo en el Primer Trimestre al Bajar el Paro en 50.000 Personas, un 10.5%. *El Mundo*. 2021. Available online: <https://www.elmundo.es/economia/2021/04/29/608a7892fc6c83406b8b4613.html> (accessed on 12 August 2021).
45. Instituto Nacional de Estadística (INE). Official Population Figures from Spanish Municipalities: Revision of the Municipal Register. Available online: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177011&menu=resultados&idp=1254734710990 (accessed on 12 August 2021).
46. Instituto Nacional de Estadística (INE). Surface Area of the Autonomous Communities and Provinces, by Altimetric Areas. Available online: <https://www.ine.es/inebaseweb/pdfDispatcher.do?td=154090&L=0> (accessed on 12 August 2021).
47. Instituto Nacional de Estadística (INE). Spanish Regional Accounts. Available online: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736167628&menu=ultiDatos&idp=1254735576581 (accessed on 12 August 2021).
48. Ministerio de Transportes Movilidad y Agenda Urbana. Tráfico en Los Aeropuertos Españoles. 2020. Available online: https://www.mitma.gob.es/recursos_mfom/listado/recursos/trafico_en_los_aeropuertos_espanoles_-_2020.pdf (accessed on 12 August 2021).
49. Sempere, P. Las 6.5 Millones de Segundas Residencias Esperan la Llegada de Sus Propietarios. *CincoDías*. 2020. Available online: https://cincodias.elpais.com/cincodias/2020/06/11/economia/1591881316_405149.html (accessed on 12 August 2021).
50. Instituto Nacional de Estadística (INE). Continuous Register Statistics. Available online: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177012&menu=ultiDatos&idp=1254734710990 (accessed on 12 August 2021).
51. CyLTV. Segovia Recupera a Sus Turistas Madrileños. *Cyltv.es*. 2021. Available online: <https://www.cyltv.es/noticia/3621F46F-D5FD-2D59-D9018F62E7B2DF86/20210509/segovia-recupera-a-sus-turistas-madrilenos> (accessed on 12 August 2021).
52. Del Campo, A. Los Madrileños Siguen Muy Presentes en Aranda del Duero. *Diario de Burgos*. 2020. Available online: <https://www.diariodeburgos.es/noticia/z5e87b0ad-0844-f501-dd699cca705bf6c3/los-madrilenos-siguen-muy-presentes-en-aranda-de-duero> (accessed on 12 August 2021).
53. Instituto Nacional de Estadística (INE). Pilot Study on Mobility Based on Mobile Phone Positioning. Available online: https://www.ine.es/experimental/movilidad/experimental_em1.htm (accessed on 12 August 2021).
54. Instituto Nacional de Estadística (INE). Residents Travel Survey. Available online: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176990&menu=ultiDatos&idp=1254735576863 (accessed on 12 August 2021).
55. ANGED. Los Efectos Socioeconómicos de la Liberalización de los Horarios Comerciales en la Comunidad de Madrid. 2017. Available online: https://www.anged.es/wp-content/uploads/2017/02/AFI_Informe-liberalizaci%C3%B3n-horarios-comerciales_comunidad_madrid.pdf (accessed on 12 August 2021).
56. Boletín Oficial del Estado (BOE). Real Decreto 926/2020, of 25 October, Declaring a State of Alarm to Contain the Spread of Infections Caused by SARS-CoV-2. 2020. Available online: <https://www.boe.es/eli/es/rd/2020/10/25/926> (accessed on 12 August 2021).
57. Boletín Oficial del Ayuntamiento de Madrid. BOAM n° 8796, 29 December 2020. Available online: https://sede.madrid.es/csvfiles/UnidadesDescentralizadas/UDCBOAM/Contenidos/Boletin/2020/Diciembre/Ficheros%20PDF/BOAM_8796_28122020143313870.pdf (accessed on 12 August 2021).
58. Boletín Oficial de la Comunidad de Madrid. BOCM n° 316, 29 December 2020. Available online: http://www.bocm.es/boletin/CM_Boletin_BOCM/2020/12/29/31600.PDF (accessed on 12 August 2021).
59. Pérez, V.; Aybar, C.; Pavía, J.M. *Dataset of the COVID-19 Lockdown Survey Conducted by GIPEyOP in Spain*; under review.
60. Kirchherr, J.; Charles, K. Enhancing the simple diversity of snowball samples: Recommendations from a research project on anti-dam movements in Southeast Asia. *PLoS ONE* **2018**, *13*, e0201710. [CrossRef]
61. Pavía, J.M.; Gil-Carceller, I.; Rubio-Mataix, A.; Coll, V.; Alvarez-Jareño, J.A.; Aybar, C.; Carrasco-Arroyo, S. The formation of aggregate expectations: Wisdom of the crowds or media influence? *Contemp. Soc. Sci.* **2019**, *14*, 132–143. [CrossRef]
62. Valliant, R.; Dever, J.A.; Kreuter, J. *Practical Tools for Designing and Weighting Survey Samples*, 2nd ed.; Springer: Cham, Switzerland, 2018.
63. Pavía, J.M.; Coll-Serrano, V.; Cuñat-Giménez, R.; Carrasco-Arroyo, S. Data from the GIPEyOP online election poll for the 2015 Spanish General election. *Data Brief* **2020**, *31*, 105719. [CrossRef]
64. Pérez, V.; Aybar, C.; Pavía, J.M. *Dataset of the COVID-19 Post-Lockdown Survey Conducted by GIPEyOP in Spain*; under review.
65. Pakpour, A.H.; Griffiths, M.D. The fear of COVID-19 and its role in preventive behaviors. *J. Concurr. Disord.* **2020**, *2*, 58–63.
66. Martínez-Lorca, M.; Martínez-Lorca, A.; Criado-Álvarez, J.J.; Cabañas Armesilla, M.D.; Latorre, J.M. The fear of COVID-19 scale: Validation in Spanish university students. *Psychiatry Res.* **2020**, *293*, 113350. [CrossRef]

67. Soraci, P.; Ferrari, A.; Abbiati, F.A.; Del Fante, E.; De Pace, R.; Urso, A.; Griffiths, M.D. Validation and psychometric evaluation of the Italian version of the fear of COVID-19 scale. *Int. J. Ment. Health Addict.* **2020**. [[CrossRef](#)] [[PubMed](#)]
68. Krishnan, V.; Gupta, R.; Grover, S.; Basu, A.; Tripathi, A.; Subramanyam, A.; Nischal, A.; Hussain, A.; Mehra, A.; Ambekar, A.; et al. Changes in sleep pattern and sleep quality during COVID-19 lockdown. *Indian J. Psychiatry* **2020**, *62*, 370–378. [[CrossRef](#)]
69. Ingram, J.; Maciejewski, G.; Hand, C.J. Changes in diet, sleep, and physical activity are associated with differences in negative mood during COVID-19 lockdown. *Front. Psychol.* **2020**, *11*, 588604. [[CrossRef](#)] [[PubMed](#)]
70. Schimmenti, A.; Billieux, J.; Starcevic, V. The four horsemen of fear: An integrated model of understanding fear experiences during the COVID-19 pandemic. *Clin. Neuropsychiatry* **2020**, *17*, 41–45. [[CrossRef](#)]
71. Block, P.; Hoffman, M.; Raabe, I.J.; Dowd, J.B.; Rahal, C.; Kashyap, R.; Mills, M.C. Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nat. Hum. Behav.* **2020**, *4*, 588–596. [[CrossRef](#)]
72. Hagerty, S.L.; Williams, L.M. The impact of COVID-19 on mental health: The interactive roles of brain biotypes and human connection. *Brain Behav. Immun. Health* **2020**, *5*, 100078. [[CrossRef](#)] [[PubMed](#)]
73. Rubin, O.; Nikolaeva, A.; Nello-Deakin, S.; Brömmelstroet, M. What Can We Learn from the COVID-19 Pandemic about How People Experience Working from Home and Commuting? Centre for Urban Studies, University of Amsterdam. 2020. Available online: <https://urbanstudies.uva.nl/content/blog-series/covid-19-pandemic-working-from-home-and-commuting.html> (accessed on 12 August 2021).
74. Thomas, F.M.; Charlton, S.G.; Lewis, I.; Nandavar, S. Commuting before and after COVID-19. *Transp. Res. Interdiscip. Perspect.* **2021**, *11*, 100423. [[CrossRef](#)] [[PubMed](#)]
75. Hensher, D.A.; Wei, E.; Beck, M.J.; Balbontin, C. The impact of COVID-19 on cost outlays for car and public transport commuting—The case of the Greater Sydney Metropolitan Area after three months of restrictions. *Transp. Policy* **2021**, *101*, 71–80. [[CrossRef](#)]
76. Nhamo, G.; Dube, K.; Chikodzi, D. Restaurants and COVID-19: A focus on sustainability and recovery pathways. In *Counting the Cost of COVID-19 on the Global Tourism Industry*; Springer: Cham, Switzerland, 2021; pp. 205–224. [[CrossRef](#)]
77. Retail Actual. La Caída del Consumo Fuera del Hogar, Cambio Radical en los Hábitos de Consumo y en Toda la Cadena de Valor. 2021. Available online: <https://www.retailactual.com/noticias/20210129/eit-food-impacto-covid-secotr-agroalimentario-europa#.YRKqEkztaUk> (accessed on 12 August 2021).
78. Madrid City Council's Open Data Portal. Censo de Locales, Sus Actividades y Terrazas de Hostelería y Restauración. Available online: <https://datos.madrid.es/sites/v/index.jsp?vgnextoid=66665cde99be2410VgnVCM1000000b205a0aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD> (accessed on 12 August 2021).
79. Somolinos, D.; Sanz, L.A. Los Hosteleros de Madrid Piden Ampliar Las 'Terrazas Covid' Para no Perder 6.000 Empleos: "Nos Han Salvado la Vida". *El Mundo*. 2021. Available online: <https://www.elmundo.es/madrid/2021/05/20/60a53a00fdddf4b8b4634.html> (accessed on 12 August 2021).
80. Casado, D. El Ayuntamiento de Madrid Retirá Todos Los Permisos Para Terrazas Sobre Aparcamientos el 1 de Enero de 2022. *elDiario.es*. 2020. Available online: https://www.eldiario.es/madrid/somos/noticias/ayuntamiento-madrid-retirara-permisos-terrazas-aparcamientos-1-enero-2022_1_7941148.html (accessed on 12 August 2021).
81. Instituto Nacional de Estadística (INE). Mean and Median Income Indicators. Available online: <https://www.ine.es/jaxiT3/Tabla.htm?t=30896&L=0> (accessed on 12 August 2021).
82. Madrid City Council Website. Se Amplía la Vigencia de las Medias Excepcionales a Las Terrazas Hasta Finales de 2021. Available online: <https://www.madrid.es/portales/munimadrid/es/Inicio/El-Ayuntamiento/Todas-las-noticias/Se-amplia-la-vigencia-de-las-medidas-excepcionales-a-las-terrazas-hasta-finales-de-2021/?vgnnextfmt=default&vgnnextoid=902f0d2685ca8710VgnVCM2000001f4a900aRCRD&vgnnextchannel=e40362215c483510VgnVCM2000001f4a900aRCRD> (accessed on 12 August 2021).
83. Madrid Es Noticia. Mahou San Miguel Triplica su Inversión en Las Terrazas de Madrid. Available online: <https://www.madridesnoticia.es/2020/05/mahou-san-miguel-triplica-su-inversion-en-las-terrazas-de-madrid/> (accessed on 12 August 2021).