

INCIDENCIAS DURANTE EL TRABAJO DE CAMPO CON ENCUESTAS PERSONALES EN ESTUDIOS EPIDEMIOLÓGICOS

Xavier Gimeno Torrent

xgimeno@ccoo.cat

ISTAS — Instituto Sindical de Trabajo, Ambiente y Salud

Centro de Referencia en Organización del Trabajo y Salud. Via Laietana, 16, 6ª planta. 08003 Barcelona.

Clara Llorens

ISTAS; Departament de Sociologia, Universitat Autònoma de Barcelona

Salvador Moncada

smoncada@ccoo.cat

ISTAS; Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra

Albert Navarro

Laboratori de Bioestadística i Epidemiologia, Universitat Autònoma de Barcelona)

Resumen: OBJETIVO: El objetivo de este artículo es describir cuáles son las incidencias más comunes que tienen lugar durante la realización del trabajo de campo en investigaciones epidemiológicas mediante encuesta personal. MÉTODO: Para ello se construyó una base de datos que contenía las incidencias registradas a lo largo de la segunda —de un total de tres— oleada (2540 encuestas) de la investigación epidemiológica «exposición a factores psicosociales en el trabajo, precariedad laboral, doble presencia y salud», a lo largo de la cual se realizaron en total 7650 entrevistas. La base de datos de incidencias contenía, además de las variables dependientes relativas a las incidencias (7 variables), otras 3 variables que servían para describir, desde el punto de vista sociodemográfico, a las secciones censales donde se realizaron las encuestas. Esta base de datos se analizó usando la técnica del análisis de tablas de contingencia mediante χ^2 , por tratarse de una técnica a medio camino entre la descripción y el análisis, muy adecuada para la investigación de objetos aún poco definidos (como es el caso). RESULTADOS: Los resultados sugieren la necesidad de considerar dos niveles explicativos: el municipio en cuanto al número de habitantes, como primera característica, y las características sociodemográficas de las secciones censales, ya que las incidencias se relacionan diferentemente con unas (por ejemplo, la inmigración en determinadas secciones) u otras características de las secciones censales (por ejemplo, el envejecimiento de la población en otras secciones) según se trate de capitales de provincia, de ciudades grandes, pequeñas, o pueblos.

Palabras clave: cuestionario Istars21 (CoPsoQ), encuestas personales, incidencias, investigación epidemiológica, municipios, secciones censales.

Abstract: OBJECTIVE: The aim of this article is to describe which are the most common incidences during data collection through personal surveys on epidemiological studies. METHOD: With this finality, a data base containing the reported incidences for wave 2's «exposición a factores psicosociales en el trabajo, precariedad laboral, doble presencia y salud» study was constructed. Number of surveys for wave 2 was 2540, the total number of waves was 3, and the total number of surveys was 7650. This data base included, leaving aside the 7 dependent variables for each type of incidence, 3 variables serving to describe, from a sociodemographic point of view, the census blocks where the surveys took place. Data base was analyzed using Crosstabs and Chi-Square Test to obtain measures of association, because this technique, half-descriptive half-analytical, is adequate to investigate unclearly-defined objects of research (and this is the case here). RESULTS: Results suggest that must be considered a two-level-approach in description of incidences —the number of inhabitants in the municipality, as the first level, and sociodemographic features of the census blocks, as the second level— because incidences are differently related to census blocks' characteristics depending on the number of inhabitants of the municipality: for example, in the provincial capital, number of immigrants are associated to certain incidence, while the same incidence are associated to another characteristic (for example, ageing of population) of the census blocks in a village.

Keywords: Istars21 (CoPsoQ) questionnaire, personal surveys, incidences, epidemiological research, municipalities, census blocks.

NOTA: Queremos agradecer muy especialmente la colaboración de CUANTER, empresa que realizó el trabajo de campo sin el cual nuestro proyecto hubiera sido imposible. También queremos agradecer al Fondo de Investigación Sanitaria del Instituto de Salud Carlos III (Ministerio de Sanidad) el inestimable apoyo que prestó desde el principio a esta investigación.

Objetivo del artículo y breve descripción del contexto general de la investigación

El principal objetivo de este artículo es *describir* empíricamente cuáles son las incidencias más comunes que tienen lugar durante la realización del trabajo de campo en investigaciones epidemiológicas mediante encuesta personal. En el estadio en que se encuentran las *escasas*, tal como se apunta en la presentación de este número monográfico de *Metodología de Encuestas* que posiblemente sea pionero en el abordaje de este tema en España, aportaciones sobre incidencias en la investigación con encuestas en nuestro país sería asumir un riesgo demasiado elevado el ofrecer una aproximación que se situara más allá de la mera descripción de las incidencias registradas para alcanzar el terreno de la interpretación y el análisis. El terreno no está maduro aún en nuestro país para alcanzar este último objetivo. Dos claros indicios de este hecho que dan, a la vez, razón de él son bastante esclarecedores: es aún relativamente poco común la práctica de documentar rigurosamente las incidencias que han acontecido durante la realización del trabajo de campo y, lo que es más importante, la de entregar sistemáticamente estas informaciones; e incluso cuando esta es una costumbre arraigada, el inventario de las incidencias se deja más al albur de cada encuestador, a causa de un menor hincapié durante el período de formación de los encuestadores —que se concentra en las peculiaridades del encargo a realizar, del cuestionario que se usará, y de la población diana que se busca— en la definición, la clasificación y el procedimiento de registro claro y completo de las incidencias producidas. Esos problemas impiden el mínimo acuerdo (entre investigadores, instituciones y empresas participantes en el campo de la investigación por encuestas) necesario para lograr una descripción válida, sistemática y completa del fenómeno de las incidencias en el trabajo de campo, requisito indispensable para alcanzar una comprensión del fenómeno mediante los diversos factores susceptibles de explicarlo. Así es como muy prudentemente —forma de obrar que debe más a la voluntad de consolidar lo poco que hay hasta ahora sobre el tema que a la carencia de valor de la propuesta del monográfico, que no creemos falta de interés vista la escasez de material en España sobre incidencias en investigación mediante encuestas...— nuestra contribución se centrará exclusivamente en la descripción empírica de las incidencias que hemos registrado durante la realización del trabajo de campo en nuestro proyecto de investigación. Y sin ninguna duda también servirá para plantear muchas más incógnitas de las que despejará.

Los objetivos principales del proyecto «exposición a factores psicosociales en el trabajo, precariedad laboral, doble presencia y salud» (Expediente PI031499, Fondo de Investigación Sanitaria del Instituto de Salud Carlos III — Ministerio de Sanidad) son tres: 1) conocer los valores de referencia de exposición a 20 dimensiones psicosociales más la dimensión doble presencia para la población ocupada española; 2) conocer la relación entre inseguridad laboral y salud; y 3) entre precariedad laboral y salud. El logro de estos objetivos puramente científicos se enmarca en la estrategia de promoción de la salud laboral que impulsa ISTAS a través, entre otras muchas acciones, de la implantación progresiva y la mejora del método Istars21 (CoPsoQ) (Moncada; Llorens; Kristensen; Vega, 2006) —denominado método PSQ CAT21 COPSOQ en Catalunya (Generalitat de Catalunya, 2005)— de evaluación e intervención sobre los riesgos psicosociales en las empresas españolas, que está disponible gratuitamente a través de la página Web de ISTAS (www.istas.net).

Para satisfacer estos diferentes objetivos la población objeto de estudio de la encuesta no podía ser otra que la formada por personas de entre 16 y 65 años, residentes en territorio nacional (se exceptuaron Ceuta y Melilla), que durante la semana de referencia hubieran estado trabajando durante al menos una hora a cambio de una retribución en dinero o en especie. También se incluyó a aquellos que teniendo trabajo hubieran estado temporalmente ausentes del mismo por enfermedad, vacaciones, etc.

El tipo de muestreo que se llevó a cabo para llegar a esta población fue el muestreo polietápico por conglomerados. Las unidades muestrales establecidas en cada etapa fueron las siguientes: municipios (1ª etapa), secciones censales (2ª etapa), hogares o viviendas familiares (3ª etapa), e individuos, como elementos muestrales finales. En tanto en cuanto la estratificación de las unidades muestrales favorece la constitución de unidades más homogéneas y el aumento de la precisión de las estimaciones se llevó a cabo este proceso, que se practicó a nivel de la primera etapa del muestreo (municipios). Se fijaron dos criterios de estratificación: la comunidad autónoma (17 estratos) y el tamaño del hábitat de residencia. Respecto al tamaño del hábitat se consideraron 5 estratos: «capitales de provincia» (estrato 1), «otros municipios de más de 50000 habitantes» (estrato 2), «municipios de 10001 a 50000 habitantes» (estrato 3), «municipios de 2001 a 10000 habitantes» (estrato 4), y «municipios de hasta 2000 habitantes» (estrato 5). La afijación muestral de cada estrato resultante del cruce entre hábitat y comunidad autónoma se llevó a cabo de forma proporcional a la población existente.

Para garantizar la representatividad de los resultados obtenidos (*validez externa*) (Cea D'Ancona, 1999: 119) se recurrió a distintas estrategias. En primer lugar, se estableció la necesidad de obtener muestras autoponderadas, que en muestreo por conglomerados en varias etapas se produce cuando las secciones censales se seleccionan con probabilidad proporcional al tamaño y las viviendas con probabilidad igual, siempre que el número de viviendas por sección sea constante e igual. En segundo lugar, se determinó que la selección de la vivienda se llevase a cabo mediante muestreo sistemático con arranque aleatorio a partir del callejero disponible para cada sección, estableciendo el punto de arranque y la cadencia de selección que garantizasen una probabilidad igual de selección para cada vivienda, y, en definitiva, para

cada individuo, en tanto que se realizó una sola entrevista por vivienda. Para ello, tanto el punto de arranque aleatorio como la secuencia del muestreo sistemático, fueron asignados a cada sección a priori de forma que el libre albedrío del entrevistador fuera nulo. Se previó también la necesidad de controlar y registrar las posibles incidencias existentes de cara a controlar y subsanar a tiempo los problemas que se presentasen durante la realización del trabajo de campo. En tercer lugar, se requirió que el proceso de selección de la persona a entrevistar dentro de cada vivienda se llevara a cabo mediante muestreo estrictamente aleatorio (mediante una tabla de números aleatorios) entre el conjunto de personas residentes en la misma que cumplieran las condiciones específicas de este estudio, además de determinados criterios correctores que ayudan a seleccionar los elementos muestrales finales de forma más adecuada. En este sentido, el grupo de edad y el sexo fueron consideradas variables auxiliares. El criterio de aleatoriedad y equiprobabilidad se respetó en todo momento, incluso más allá del muestreo, en el momento de efectuar las sustituciones, ya que no hacerlo de esta forma supondría quebrar la regla de aleatoriedad y equiprobabilidad. Es por esto que las sustituciones, ya fueran de individuos, de hogares, o de secciones censales, se realizaron observando reglas fijas fijadas de antemano.

Por otra parte, y en cuanto a otra de las fuentes tradicionales de error, cabe decir que el cuestionario utilizado fue ya validado en su momento. La *validez de constructo* (Cea D'Ancona, 1999: 119-120) del instrumento de medida usado, el cuestionario Ista21 (CoPsoQ), fue comprobada mediante el método de consistencia interna alfa de Cronbach, y sus resultados son los siguientes: la alfa de Cronbach osciló entre 0,65 y 0,92, excepto en dos casos a causa del escaso número de ítems que contenían las escalas. Esto significa que se reprodujeron de forma favorable las asociaciones esperadas para el marco conceptual usado, que las medidas de los errores cometidos eran aceptables y, por tanto, que la fiabilidad del instrumento de medida era la adecuada (Moncada; Llorens; Navarro; Kristensen, 2005). La duración del cuestionario se fijó en entre 40 y 45 minutos, y estaba compuesto por preguntas cerradas.

Para garantizar la máxima fiabilidad de las estimaciones realizadas (*validez estadística o de conclusión estadística*) (Cea D'Ancona, 1999: 120-121) se requirió la necesidad de disponer de tamaños muestrales adecuados en los diferentes estratos de la muestra. El criterio más determinante en este sentido es el tamaño muestral, ya que determina la posibilidad posterior de que los diferentes estratos muestrales tengan un tamaño muestral adecuado. Se fijó, así, un tamaño muestral de 7650 individuos, lo que supone un error muestral máximo del $\pm 1,14$ en las siguientes condiciones: $P=Q=50\%$ y para un nivel de confianza del $95,5\%$ (2σ). Esto supone la consecución de una muestra adecuada, la ausencia de sesgo de la cual fue también comprobada. Los 7650 individuos fueron entrevistados por oleadas, de forma que en la primera oleada se entrevistó a 2560 personas, en la segunda a 2540 y en la tercera a 2550. En este sentido, el número de secciones censales que fueron extraídas aleatoriamente de cara a la cumplimentación del número de encuestas a realizar en cada oleada fue de 256, y siempre fueron las mismas; es decir, fueron escogidas de una vez por todas y se visitaron en cada una de las tres oleadas, cambiándose únicamente las posibles zonas de encuesta en cada oleada para que no pudiera entrevistarse más de una vez a una misma persona. En cada una de las secciones se realizaron 10 entrevistas.

Los resultados que se presentan en este artículo se refieren *únicamente* a las 254 secciones censales (correspondientes a 2540 entrevistas) que fueron visitadas durante la realización del trabajo de campo que constituyó la segunda oleada, que tuvo lugar del 08/02/2005 al 08/03/2005. Se ha optado por esta alternativa por tres razones. La primera ya que la oleada es la mínima unidad temporal de muestreo, y añadir más de una oleada supondría sumar diferentes momentos de encuesta, con sus incidencias consiguientes sumadas artificialmente, ya que en ningún momento han sido producto de un mismo momento de encuesta, y este hecho podría suponer la introducción de sesgos considerables. La segunda ya que sumar más de una oleada supondría algo similar a duplicar secciones censales (algo así como si se encuestase dos veces a la misma persona), cosa que implicaría la duplicación de las incidencias, cosa que en ningún caso es cierta. La tercera ya que sumar más de una oleada significaría, bajo el supuesto de alternancia de zonas censales (calles) dentro de una misma sección censal en las diferentes oleadas, la imposibilidad de relacionar las incidencias que tienen lugar en cada una de las oleadas con las características sociodemográficas de cada una de las zonas censales que son susceptibles de explicarlas: si incidencias que se producen en zonas distintas están mezcladas y las características sociodemográficas de cada una de estas zonas tampoco son distinguibles, ¿cómo relacionar las incidencias con estas características?

Método seguido y presentación de los resultados

El registro de las incidencias producidas durante la realización del trabajo de campo es una parte fundamental de la labor destinada a garantizar el cumplimiento de la secuencia de muestreo sistemático que ha sido planificada de antemano, y que acaba de ser descrita para nuestra investigación. Una herramienta fundamental en el desempeño de esta tarea son las hojas de ruta. El registro de estas incidencias no es necesario sólo a efectos de control, sino que también es indispensable de cara a poder gestionar los errores que no tienen que ver ni con el muestreo, ni con el cuestionario, ni con el profesional que encuesta: se trata de los errores derivados de las sustituciones realizadas.

A través de las hojas de ruta se obtuvo la información relativa a las incidencias, que fue volcada en la base de datos de incidencias que se construyó a tal efecto dando lugar a 7 variables *dependientes* —esta información además de toda la que se comenta en este apartado aparece en los anexos que acompañan a este artículo—: «No contactos de hogar», «Negativas individuales», «Hogares sin personas localizables», «Ausencias de persona seleccionada en hogar», «Imposibilidad por idioma», «Otro tipo de imposibilidad» y «Población no encuestable». Los no contactos de hogar se

refieren tanto a la imposibilidad de acceder al hogar (por ejemplo, porqué en algunas zonas no se permite el acceso a determinados edificios a que la ruta aleatoria nos dirigía, ya que puede haber conserjes, porteros, vigilancia, etc.) como a la inexistencia del supuesto hogar (por ejemplo, porqué en lugar de una vivienda familiar la ruta aleatoria fijada conduce a un local de oficinas, a un establecimiento comercial, etc.). Las negativas individuales consisten en el hecho que, una vez se ha accedido al hogar, sus habitantes se niegan a participar en el estudio. Los hogares sin personas localizables son aquellas viviendas familiares en que el profesional que realiza la encuesta no logra encontrar a nadie por las razones que fueran. En este caso, el planteamiento de nuestra encuesta consistió en no sustituir el hogar hasta después de un mínimo de tres visitas al mismo en horarios y días alternos (incluyendo fines de semana). La ausencia de persona seleccionada en hogar se refiere al hecho que se accede al hogar y se localiza a personas que acceden a participar en la encuesta, pero la persona que se requiere para participar, ya que es la que cumple los requisitos fijados, no se encuentra disponible en la vivienda en ese momento. En este caso, en nuestra investigación, se obró de forma que se intentó concertar cita con la persona y, por tanto, no se efectuó su sustitución hasta que no existiera negativa expresa del individuo o hasta que el número de citas concertadas sin acudir fuera superior a dos. La imposibilidad por idioma representa la imposibilidad de comprender o expresarse adecuadamente en castellano, y, por tanto, la imposibilidad de ser encuestado y responder la encuesta. La denominación «Otro tipo de imposibilidad» es suficientemente explícita, pero también muy ambigua. La población no encuestable se refiere a la población que no cumplía los criterios de selección de las personas que podían ser susceptibles de participar en el estudio. En la codificación de estas 7 variables dependientes se decidió considerar 2 únicas categorías en las que la frecuencia de cada una de las incidencias no jugaba ningún papel: «Sin incidencias» y «Con incidencias», ya que se consideró que en el estudio de las incidencias un primer estadio de investigación debía ser la descripción de las incidencias y las no incidencias para, posteriormente, cuando se hubieran comprendido estos dos fenómenos empezar a estudiar, a través de análisis más complejos, la frecuencia de cada uno de estos fenómenos.

Además de estas 7 variables dependientes, en la base de datos de incidencias también se incluyeron 6 variables adicionales, que recibieron dos tratamientos diferenciados. Así, 3 de estas variables fueron consideradas variables *intermedias*, y se referían a una característica como el número de habitantes del municipio. Se trataba del «Tipo de municipio» —se dividió a los municipios según fueran capitales de provincia o no capitales— y de la categorización de los distintos municipios por estratos de acuerdo con el número de habitantes en el municipio. Se practicaron dos categorizaciones diferentes según el número de estratos fijados, que o bien podía ser de 5 —capitales de provincia, otros municipios de más de 50000 habitantes, municipios de 10001 a 50000 habitantes, municipios de 2001 a 10000 habitantes, y municipios de hasta 2000 habitantes— o de 2 —municipios de hasta 2000 habitantes y municipios de 2001 y más habitantes—. Por otra parte, 3 de estas 6 variables fueron consideradas propiamente variables *independientes*, que se referían a las características sociodemográficas de las diferentes secciones censales visitadas. Se tomaron estas 3 variables independientes ya que eran las únicas accesibles de forma pública y gratuita a través del directorio Web de INE (www.ine.es). Se trata de las siguientes variables: «proporción de mayores de 65 años en sección censal», «proporción de inmigrantes en sección censal» y «proporción de ocupados de 16 años o más en sección censal». Estas variables independientes fueron sometidas a una categorización por terciles. Se consideró esta codificación la más pertinente para someter a análisis de tablas de contingencia las asociaciones establecidas entre los 3 tipos de variables que acaban de ser apuntados.

Para estudiar estas asociaciones, y teniendo en cuenta que se trataba de un objeto de estudio que se encontraba en plena fase de definición en nuestro país —y es muy posible que las peculiaridades nacionales en cada uno de los distintos ámbitos sociales en que se centran las distintas investigaciones (en nuestro caso, especialmente, el mercado de trabajo) pesen mucho en la definición de las incidencias que tienen lugar en las investigaciones por encuesta que acontecen en cada país— y que, por tanto, no era posible apuntar aún ni hipótesis claras ni un modelo analítico que ayudara a analizar y a interpretar el fenómeno de las incidencias, se consideró que la técnica más adecuada a emplear era el análisis de tablas de contingencia mediante la prueba de χ^2 . Se adoptó como nivel de significación a partir del cual las asociaciones entre variables serían señaladas como relevantes el umbral de 0,10 ($\alpha=0,10$ o nivel de confianza del 90%). Así es como todas las asociaciones entre variables que se han plasmado en los anexos correspondientes en que se basan todos los análisis apuntados en esta sección se caracterizan (casi siempre) por ser significativas para un NC=90%. Sólo en 4 casos esta regla es quebrantada.

PRIMER NIVEL DE ANÁLISIS (análisis bivariado de tablas de contingencia). En el primer nivel de análisis considerado se sometieron a la prueba de χ^2 las asociaciones establecidas entre todas las variables dependientes consideradas y, alternativamente, las variables denominadas intermedias (caracterización de los municipios según el número de habitantes) y, en segundo lugar, las 3 variables independientes consideradas (caracterización sociodemográfica de las secciones censales).

En cuanto a las variables intermedias, se establece, en primer lugar, una asociación poco intensa (V Cramer=0,100), que limita el nivel de significación considerado apropiado ($p=0,110$), entre el tipo de municipio y las negativas individuales, de manera que las capitales de provincia se asocian con este tipo de incidencias, cosa que en las no capitales no sucede. También se aprecia una asociación, en esta ocasión muy significativa ($p=0,016$) pero poco intensa (V Cramer=0,151), entre, de nuevo, el tipo de municipio y las ausencias de persona seleccionada en el hogar,

que sigue la anterior pauta: en las capitales de provincia se producen incidencias mientras que en las no capitales no. A pesar de que este primer análisis de las asociaciones establecidas entre variables intermedias y variables dependientes es francamente poco esclarecedor, pone sobre aviso de una pauta que se repite en los dos casos: la capitalidad de un municipio se asocia de forma significativa, aunque ciertamente poco intensa, con el hecho de tener incidencias de estos dos tipos (negativas individuales y ausencias de persona seleccionada en hogar) cuando se realizan encuestas personales. Esta misma pauta puede observarse desde un punto de vista un poco distinto, a la luz del análisis de las asociaciones establecidas entre la segunda variable intermedia considerada y las distintas variables dependientes. En este sentido se establecen asociaciones entre cada uno de los dos tipos de municipios considerados según el número de habitantes «categorización por estratos (2 estratos)» y distintas variables dependientes, de forma que los municipios de hasta 2000 habitantes se asocian de forma muy significativa, y casi igual de intensa que en los anteriores casos, con no tener incidencias relativas a los no contactos de hogar, las negativas individuales y determinadas imposibilidades de ser encuestado (respectivamente: $p=0,013$, $p=0,024$, $p=0,035$). Contrariamente, los municipios de 2001 y más habitantes — entre los cuales se encuentran las capitales de provincia— se asocian significativamente con tener este tipo de incidencias. Pero estas asociaciones sólo se observan si se toman en bloque todos los municipios de 2001 y más habitantes. Esta observación conduce tanto a una crítica como a una constatación: considerar, de forma relativamente artificial, un monolito a municipios tan diversos no es aconsejable si se pretende entender cada caso; aunque también es cierto que, según parecen señalar los análisis efectuados hasta ahora, el umbral de los 2000 habitantes debe ser necesariamente considerado si se pretende distinguir entre aquellos municipios en que se presentan incidencias más frecuentemente y aquellos en que se presentan de forma más residual. Esta valoración debe tenerse en cuenta, ya que, como señalará el análisis de tablas de contingencia trivariado, tanto el umbral de los 2000 habitantes como la capitalidad de las provincias son los dos extremos que, básicamente, se asocian con el hecho de no tener o tener incidencias. Pero sin embargo, tomado en su conjunto, el análisis recién comentado no puede arrojar otra conclusión que el hecho de que las asociaciones establecidas entre las variables dependientes y las variables intermedias consideradas en este primer nivel de análisis arrojan unos resultados bastante discretos.

Siguiendo aún en este primer nivel de análisis, se procedió, en segundo lugar, a testar las asociaciones entre las 3 variables independientes consideradas y las 7 variables dependientes. La asociación más significativa ($p=0,007$) y también la más intensa de todas las que se examinaron a lo largo de este primer nivel de análisis ($V \text{ Cramer}=0,203$) es la que se establece entre la proporción de personas ocupadas y las ausencias de persona seleccionada, de forma que en aquellas secciones censales en que la proporción de personas ocupadas oscila entre el 22,29% y el 37,60% (el tercil de menor ocupación) esta incidencia no se presenta. Contrariamente, en las secciones con una proporción de ocupación más alta, entre el 37,61% y el 43,52% (tercil que acumula entre el 33% y el 66% de los casos; es decir, el tercil medio), esta incidencia suele presentarse de forma muy significativa. Por otra parte, la proporción de ocupación es una variable que, en determinados casos, parece asociarse de forma muy significativa ($p=0,036$) pero de forma poco intensa ($V \text{ Cramer}=0,166$) con las «imposibilidades por idioma», de forma que se producen incidencias en este sentido en aquellas secciones censales en que la proporción de ocupados es mayor (tercil superior: del 43,53% al 62,50% de ocupados). Es muy posible que esta asociación se dé de esta forma en aquellas secciones censales en que la proporción de inmigrantes sea relativamente alta y en que estos inmigrantes trabajen en la misma sección censal (tiendas de productos alimenticios autóctonos de las zonas de procedencia de los inmigrantes, locutorios, etc.), cosa que suele pasar en determinadas secciones, convertidas en auténticos «centros comerciales» especializados en el abastecimiento de todo tipo de productos y servicios que son regentados por los mismos inmigrantes y que se dirigen a los mismos inmigrantes. De todas formas, esta hipótesis parecería ser válida sólo en el caso de las capitales y en las grandes ciudades, ya que esta es la caracterización del tipo de municipio que subyace implícitamente en la hipótesis señalada. Este razonamiento podría explicar, por tanto, la asociación que se establece entre la proporción de ocupación y el hecho de que aparezcan como incidencia las «imposibilidades por idioma», que se hipotiza que se daría en el caso de las capitales y en otras ciudades de más de 50000 habitantes, y que no vuelve a aparecer en el tercer nivel de análisis. Se trata de una explicación relativamente razonable, ya que el análisis de la correlación establecida entre las variables % de ocupados y % de inmigrantes en sección censal, limitado únicamente a los municipios del estrato 1 y 2 —este análisis no se recoge en los anexos—, da como resultado un coeficiente de correlación de 0,524 (que es significativo al nivel 0,01). Se trata de una correlación relativamente alta que, además, es la más importante de todas las observadas entre pares de variables independientes para el caso de los municipios de estos dos estratos. Y en tanto que este primer análisis no agota ni pretende agotar la explicación que sería necesario dar a este fenómeno, sería muy aconsejable iniciar el estudio de la hipótesis explicativa que acaba de proponerse considerando las técnicas de análisis, las variables y las relaciones entre variables que no han sido consideradas en nuestro análisis.

Pero en algunos casos incluso las propuestas de explicaciones de las asociaciones constatadas mediante los tests de asociación entre variables practicados se vuelven imposibles. Tal es el caso de las asociaciones que se establecen entre, por una parte, la proporción de mayores en la sección censal y las incidencias «otro tipo de imposibilidad» y «población no encuestable»; y, por otra, la proporción de personas ocupadas y la incidencia «población no encuestable». Tal es así ya que, si bien en el caso de la proporción de mayores en sección censal y de las incidencias «otro tipo de imposibilidad» y «población no encuestable», parece ser muy razonable la constatación de asociaciones entre estas variables, asociaciones que son significativas para un nivel de confianza del 90% (respectivamente $p=0,106$ y $p=0,084$) —parece totalmente razonable que en las secciones censales el grado de envejecimiento se relacione con presentar dificultades al responder una encuesta y también, en el caso de nuestra investigación especialmente, ya que se dirige a

población en edad de trabajar, con el hecho de no formar parte de la población que se busca para formar parte del estudio—, no parece ser muy razonable que las asociaciones entre categorías de variables dependientes e independiente en este caso sean, como son, las contrarias: cuando la proporción de mayores es mayor (del 13,14% al 47,97% en el caso de la incidencia «otro tipo de imposibilidad»; del 19,87% al 47,97% en el caso de la incidencia «población no encuestable») no se presentan estas incidencias y cuando es menor (del 2,60% al 13,13% en el caso de la incidencia «otro tipo de imposibilidad»; del 13,13% al 19,86% en el caso de la incidencia «población no encuestable») se presentan. Y lo mismo ocurre en el caso de la asociación entre la proporción de población ocupada y la incidencia «población no encuestable». La constatación de esta asociación (significativa para un nivel de confianza del 90%, ya que $p=0,086$) entre variables es relativamente sorprendente. Pero lo que resulta aún más sorprendente es descubrir que el tercil superior de ocupación (que concentra a aquellas secciones censales en que la proporción de ocupación oscila entre el 43,53% y el 62,50%) se asocia muy significativamente (el residuo estandarizado 2,0 es significativo al 95%) con presentar la incidencia «población no encuestable», mientras que el tercil medio de ocupación (y, por tanto, aquellas secciones que se sitúan en este tercil) se asocia con no presentar esta incidencia. ¿Cómo explicar la anterior asociación global y el fenómeno, totalmente paradójico, consistente en que en las secciones censales con mayor ocupación la población no encuestable es comparativamente más frecuente que en las secciones censales con menor ocupación? La respuesta a esta pregunta posiblemente requerirá, tal como anteriormente se ha apuntado en el caso de la asociación establecida entre las imposibilidades por idioma y la proporción de ocupados en la sección censal, un análisis de las relaciones entre las diversas variables independientes que han sido consideradas. Pero en este caso se hacen necesarias dos medidas adicionales no apuntadas hasta ahora: la obligación de buscar nuevas variables que pudieran funcionar como variables independientes explicativas de las incidencias; y, en segundo lugar, el requerimiento de investigar cuáles son las relaciones que se establecen entre todas las variables independientes consideradas (las que nosotros hemos considerado y las que podrían considerarse) y una variable que a lo largo de estos dos casos problemáticos considerados se ha revelado tan determinante: la ocupación; variable que, además, es central en nuestro estudio, ya que se trata de encuestar a personas ocupadas. Estas dos medidas son absolutamente necesarias ya que los análisis practicados hasta ahora parecen revelar —pero esta afirmación debe tomarse con mucha cautela, ya que se trata de una hipótesis— relaciones espurias entre la proporción de ocupación y otras variables explicativas, algunas conocidas y otras desconocidas por nosotros.

SEGUNDO NIVEL DE ANÁLISIS (análisis bivariado de las asociaciones establecidas entre las variables intermedias e independientes). Con vistas a avanzar en el análisis de las asociaciones establecidas entre las variables consideradas en este artículo y, especialmente, a consecuencia de la poca potencia explicativa de las variables intermedias e independientes por separado —como acaba de ser apuntado— se efectuó el análisis de las asociaciones que mostraban las variables intermedias con las independientes. En estos análisis sólo se consideró una variable intermedia («categorización por estratos» a nivel de 5 grupos) con todas la demás variables independientes. Se procedió de esta forma ya que del total de tres variables intermedias construidas, una de ellas («tipo de municipio») presentaba una alta correlación (el coeficiente de correlación de Pearson es $-0,791$, significativo al nivel 0,01) con la variable «categorización por estratos (5 estratos)», y la segunda «categorización por estratos (2 estratos)» se trataba de una categorización distinta de la variable «categorización por estratos».

Las asociaciones analizadas entre la variable intermedia y las 3 variables independientes consideradas fueron siempre muy significativas (el p-valor fue siempre igual o inferior a 0,001), excepto en el caso de la proporción de personas ocupadas, que no se asociaba de forma significativa con la variable «categorización por estratos (a 5 estratos)». La intensidad de estas asociaciones ha sido mejor que la que se analizó en el anterior apartado (V Cramer oscilaba en el anterior nivel entre 0,100 y 0,203), ya que ha oscilado entre $V\text{ Cramer}=0,348$ y $V\text{ Cramer}=0,425$.

Se ha observado asociación significativa entre la variable intermedia y las variables independientes «% de mayores de 65 años» y «% de inmigrantes». En el primer caso, los municipios de más de 50000 habitantes y los que tienen entre 10001 y 50000 habitantes se caracterizan por una baja proporción de personas mayores en sus secciones censales —la más baja de todas, de entre el 2,60% y el 13,13%—, contrariamente los que tienen entre 2001 y 10000 habitantes se distinguen por tener una proporción entre media y elevada de gente mayor, ya que se sitúan indistintamente en los terciles medio —entre el 13,14% y el 19,86% de mayores de 65 años— y superior —entre el 19,87% de mayores de 65 años y el 47,97%—. Por último, las secciones censales de las capitales de provincia y de los municipios de hasta 2000 habitantes son los que tienen una mayor proporción de personas mayores de 65 años, en tanto que se sitúan en el tercil superior, que agrupa a aquellas secciones que tienen entre el 19,87% y el 47,97% de personas mayores de 65 años. En el caso de la última variable, tanto los municipios de 2001 a 10000 habitantes como los que tienen de 10001 a 50000 habitantes se caracterizan por tener las mayores proporciones de inmigrantes, ya que se asocian con el tercil superior (tener entre el 10,81% y el 48,91% de inmigrantes), mientras que las capitales de provincia se caracterizan por su asociación con el tercil medio de inmigración (tener entre el 6,56% y el 10,80% de inmigrantes).

TERCER NIVEL DE ANÁLISIS (análisis trivariado de tablas de contingencia: 1 variable dependiente y 1 variable independiente según cada categoría de la variable intermedia considerada «categorización por estratos (5)»). El análisis precedente auguraba una buena cosecha para el análisis de tablas de contingencia trivariado, y se practicó como consecuencia de la sospecha que el efecto explicativo de las variables independientes no podía desligarse del efecto explicativo ejercido por las variables intermedias, de forma que la fuerza explicativa de una misma variable

independiente no podía ser la misma en cada uno de los 5 tipos de municipios considerados. Y precisamente eso es lo que confirmaban los análisis practicados en el anterior nivel de análisis. De todos modos, como se observará, se pueden apreciar algunas asociaciones reiteradas entre unas mismas variables independientes y unas mismas incidencias para algunos de los estratos considerados.

A grandes rasgos, los resultados de estos análisis ratifican lo que apuntaban los análisis practicados a lo largo del primer nivel de análisis. Resaltan y confirman la importancia del umbral de los 2000 habitantes para explicar el efecto del tamaño del municipio sobre la posibilidad o no de que se produzcan incidencias durante la ejecución del trabajo de campo: de las 16 asociaciones significativas entre variables detectadas mediante el análisis trivariado, todas se producían en aquellos municipios de 2001 o más habitantes, tal como el análisis bivariado efectuado durante la primera fase del análisis avanzaba. Es decir, en este nivel de análisis no se han observado asociaciones entre incidencias y variables explicativas para los municipios de hasta 2000 habitantes.

Cabe remarcar, de todas formas, que esto no significa que no se produzcan incidencias en los municipios de hasta 2000 habitantes. Significa únicamente que para aquellas asociaciones entre algún tipo de incidencia y algunas variables independientes para cada tipo de municipio que los análisis han revelado como significativas, en los municipios de hasta 2000 habitantes no suelen presentarse incidencias. Es decir, para algún tipo de incidencias (las que no aparecen en los anexos; que son las que no han sido significativas) el umbral de los 2000 habitantes no marca ningún tipo de diferencia entre municipios, ya que no hay diferencias entre los municipios de hasta 2000 habitantes y todos los demás respecto a la observación de este tipo de incidencias. Pero en tanto que en este trabajo la estrategia no es tanto empezar a *entender lo que es igual* en todos los municipios sino *empezar a entender lo que es distinto* (es decir entender lo «significativamente distinto»; en otras palabras, empezar a describir el fenómeno desde la diferencia y no desde la similitud), hemos considerado más apropiado señalar las dinámicas diferenciales y obviar las similares.

Las principales asociaciones que el análisis trivariado revela para el caso de las **capitales de provincia** son las siguientes. En primer lugar la asociación que liga la proporción de inmigrantes en la sección censal con la no localización de personas en el hogar; y de las asociaciones entre la proporción de personas ocupadas y, por una parte, la no localización de personas en el hogar, y, por otra, la ausencia de la persona seleccionada. En el caso de la asociación, que es muy significativa ($p=0,056$) y relativamente intensa (V Cramer= $0,250$), entre la variable independiente relativa al número de inmigrantes y la no localización de personas en el hogar, las asociaciones locales entre categorías de variables son bastante claras: el tercil inferior (del 0,51% al 6,55% de inmigrantes) se asocia con no tener este tipo de incidencias, que se presenta en el caso de los terciles medio y superior (de forma conjunta; es decir, del 6,56% al 48,91%). Parece relativamente claro que en el caso de secciones censales conformadas mayoritariamente (del 6,56% al 48,91%) por inmigrantes la población autóctona suele ser percibida como distinta y, muy a menudo, como extraña al grupo y también muy a menudo como posible amenaza, ya que en muchas ocasiones la población autóctona que visita el barrio suele ser de asistentes sociales y policías; encargados, en definitiva, de «vigilar y sancionar» a los inmigrantes, y así es como los inmigrantes es posible que los perciban. Esto puede implicar el hecho que la situación de encuesta —similar al interrogatorio policial— y la figura del encuestador —posiblemente considerado una especie de policía que controla que todo esté bien a través de su cuestionario— despierten los recelos de los habitantes, que se protegen de su influjo «no abriéndole la puerta». Esta hipótesis parece bastante razonable, ya que la asociación entre esta incidencia u otras muy similares —en el caso de las no capitales la incidencia es «no contactos», y pudiera ser posible que las dos incidencias hayan sido confundidas a la hora de ser registradas— se repite de forma reiterada para los «otros municipios de más de 50000 habitantes», para los «municipios de 10001 a 50000 habitantes» y para los «municipios de 2001 a 10000 habitantes». De todas formas, sería necesario investigar más adecuadamente esta posible explicación.

Respecto a las asociaciones que ligan el número relativo de personas ocupadas con la no localización de personas en el hogar y con la ausencia de la persona seleccionada, cabe decir que, una vez más, se trata de asociaciones muy significativas (siempre para un $NC=90\%$) e igual de intensas que en el anterior caso (V Cramer= $0,251$). La asociación entre la ocupación y la no localización de personas en el hogar se presenta cuando las secciones censales se sitúan en el tercil superior de ocupación (del 43,53% al 62,50%); y cuando la ocupación es baja o media (del 22,29% al 43,52%), no se presenta esta incidencia. La pauta parece ser relativamente coherente: a mayor ocupación menos fácil será encontrar a alguien en casa, si se visitan las casas en horarios de trabajo —factor que seguramente influya muchísimo y que no hemos podido tener en cuenta, ya que en nuestras hojas de ruta no se encontraba disponible, en suficiente cuantía, información sobre los horarios de visita de los profesionales encuestadores—, y más frecuentemente se presentará esta incidencia en esas secciones, a causa del mayor número de personas ocupadas. El patrón sería el contrario (no se presentaría esta incidencia) en el caso de las secciones con menos personas ocupadas, a causa, básicamente, del menor número de personas ocupadas, y no de que estuvieran más en casa (aunque para poder verificar esta última afirmación deberían estudiarse las interacciones entre el porcentaje de ocupación y el número de horas en el trabajo, ya que pudiera darse el caso que en aquellas secciones censales donde hay más ocupación también se trabajen más horas, y a la inversa para las secciones censales con menor ocupación). En este caso es posible que se presentaran otras incidencias que vendrían explicadas por otras variables. De todos modos, estas son hipótesis que deberían ser verificadas. La asociación entre la ocupación y la ausencia de la persona seleccionada ($p=0,01$; V Cramer= $0,316$) sigue un modelo similar: las incidencias se concentran entre el tercil superior (del 43,53% al 62,50% de ocupados) y el medio (del 37,61% al 43,52% de ocupados). De hecho, es este último el que se asocia significativamente con este tipo de incidente, mientras que el tercil superior se asocia con no tener esta incidencia. Es posible que este fenómeno contraintuitivo —al parecer lo más coherente sería que a mayor tasa de ocupación (tal como ocurría en el caso anterior), mayores incidencias— tenga

que ver con dinámicas de relación entre la variable dependiente y la independiente que no sean lineales. De lo que se trataría es de estudiar cuáles son estas dinámicas de relación, y si tienen que ver con relaciones espurias establecidas con otras variables que no han sido consideradas en nuestros análisis. Este objetivo escapa totalmente de los propósitos de este trabajo, que únicamente pretende empezar a plantear algunas cuestiones que, como en este caso, deberían resolverse ulteriormente. La investigación de este tipo de dinámicas relacionales creemos que es necesaria, ya que se reproducen frecuentemente circunstancias muy parecidas en las relaciones entre las variables dependientes y las independientes, como se verá mediante diversos ejemplos a lo largo de este nivel de análisis. Sin ir más lejos, vuelve a ocurrir en la última relación significativa ($p=0,088$; V Cramer= $0,230$) entre variables que tiene lugar en el caso de las capitales de provincia. Se trata de la relación que se establece entre la proporción de mayores de 65 años y la incidencia «población no encuestable». Es bastante claro que la relación es muy coherente, ya que nuestra investigación se dirigía a personas en edad de trabajar y, en la mayor parte de los casos, tener más de 65 años significa dejar de hacerlo y, así, dejar de ser población diana en nuestra encuesta. Lo que no resulta comprensible son las causas (o las dinámicas de relación entre variables, que deberían investigarse) que explicarían el hecho que esta incidencia se produzca para el tercil medio (del 13,14% al 19,86% de mayores), pero no para el superior (del 19,87% al 47,97% de personas mayores en sección censal).

En el caso de los **otros municipios de más de 50000 habitantes** se observan las siguientes asociaciones. En primer lugar, la que liga la proporción de personas ocupadas con la ausencia de persona seleccionada en hogar ($p=0,016$; V Cramer= $0,444$). La asociación que se establece entre la ocupación y la ausencia de la persona seleccionada en el hogar es totalmente coherente con lo que la intuición nos sugeriría: esta incidencia se asocia con el tercil jerárquicamente superior (en este caso los terciles medio y superior) y no se presenta en el tercil jerárquicamente inferior (en este caso el tercil inferior). En segundo lugar, cabe señalar la relación de asociación que también se establece en este estrato entre la tasa de inmigrantes y la ausencia de la persona seleccionada en el hogar ($p=0,018$; V Cramer= $0,439$). Se trata claramente de una relación espuria entre variables en tanto que la población diana en nuestra investigación es la población ocupada, y las ausencias de la persona seleccionada se refieren a personas ocupadas y no a personas inmigrantes. Esto significa, por tanto, que la relación entre estas dos variables estaría mediatizada por la variable central en nuestro estudio (la ocupación), y se referiría a una categoría de ocupados especial: los ocupados que son inmigrantes. Y, ciertamente, las asociaciones establecidas entre la inmigración y la ausencia de la persona seleccionada parecen traducir las pautas de asociación que caracterizan a la asociación que se establece entre la ocupación y esta misma incidencia: en el caso de las dos asociaciones es el tercil inferior el que se relaciona con no tener esta incidencia y son los terciles medio y superior los que se relacionan con tenerla. Pero está claro que, como se ha repetido insistentemente, estas pautas de asociación deberían ser estudiadas, ya que pudiera darse muy bien el caso que la interacción entre estas dos variables independientes (la calidad de persona «ocupada» y de persona «inmigrante») implique una mayor frecuencia de la incidencia «ausencia de la persona seleccionada» para las «personas ocupadas» que son «inmigrantes».

Además de las comentadas, en este estrato se observan otras relaciones entre variables dependientes e independientes. Se trata de las relaciones que ligan el hecho de no contactar con hogares (ya que la ruta aleatoria ha acabado por conducir a establecimientos comerciales, oficinas o casas desabitadas, por ejemplo) con la proporción de inmigrantes ($p=0,105$; V Cramer= $0,327$); y la proporción de personas ocupadas con los hogares sin personas localizables ($p=0,089$; V Cramer= $0,339$). De nuevo, estas asociaciones, aunque globalmente muy razonables, se producen para categorías de variables que son difícilmente comprensibles. Sería necesario estudiar la forma que toman las relaciones locales entre estas variables, la causa de estos hechos, y, muy especialmente, si estas relaciones son explicativas de algo.

En cuanto a los **municipios de 10001 a 50000 habitantes**, la explicación de las asociaciones parece no ser tan conflictiva. La pauta de asociación entre la proporción de personas ocupadas y los hogares sin personas localizables ($p=0,042$; V Cramer= $0,319$) reproduce la que se producía para las capitales de provincia: es decir, se producen estas incidencias para el tercil jerárquicamente superior (terciles medio y superior) pero no para el jerárquicamente inferior (tercil inferior). Es decir, nuevamente, parece ser que cuanto mayor es la proporción de personas ocupadas más difícil es localizar a personas que estén en casa: es muy posible que —aquí debería tenerse en cuenta, de nuevo, la hora de visita— estén trabajando. Este mismo patrón explica la asociación que se establece, puede que por razones muy parecidas, entre las negativas individuales y la proporción de personas ocupadas en este estrato ($p=0,076$; V Cramer= $0,288$): el tercil medio se asocia con la participación mientras que el superior se asocia con las negativas a participar («ahora no puedo, tengo trabajo»). Por otra parte, se aprecia también una asociación que ya había aparecido en el caso de los municipios de más de 50000 habitantes, pero bajo una nueva forma: en los municipios de 10001 a 50000 habitantes las secciones censales que tienen entre el 10,81% y el 48,91% (tercil superior) de inmigrantes se asocian con tener no contactos de hogar, mientras que las que tienen entre el 6,56% y el 10,80% de inmigrantes se asocian con no tenerlos ($p=0,079$; V Cramer= $0,286$). En la explicación de esta asociación debe tenerse en cuenta la asociación que, durante el segundo nivel de análisis, se ha verificado entre la proporción de personas inmigradas y el número de habitantes: sabemos que las secciones censales pertenecientes a este estrato se asocian con tener entre el 10,81% y el 48,91% de inmigrantes; por eso es muy razonable que se observe esta incidencia para esta tasa de inmigrantes en los municipios de este estrato. En último lugar, aparece una asociación bastante inexplicable: la que se establece entre las secciones que tienen un alto porcentaje de ocupación (entre el 43,53% y el 62,50%) y la incidencia «población no encuestable», que

no se da en aquellas secciones que se caracterizan por tener un porcentaje de ocupación entre el 22,29% y el 43,52% ($p=0,110$; $V \text{ Cramer}=0,267$).

En el estrato que agrupa a los **municipios de 2001 a 10000 habitantes** se establecen 4 asociaciones significativas entre variables. La primera es la que liga, como en el anterior estrato y en su predecesor, la proporción de inmigrantes con los no contactos ($p=0,023$; $V \text{ Cramer}=0,520$), aunque en este caso es el tercil jerárquicamente inferior (se trata del tercil medio) quién se asocia con esta incidencia, que no se presenta para el tercil jerárquicamente superior (tercil superior). La segunda es la asociación que se establece entre la proporción de personas mayores de 65 años y las negativas individuales ($p=0,031$; $V \text{ Cramer}=0,499$), de forma que en el tercil jerárquicamente superior (el superior propiamente) se da esta incidencia, que no se produce en los terciles inferior y medio. La tercera asociación significativa es la que se establece entre la proporción de personas ocupadas y, nuevamente —como ocurría en el anterior estrato—, las negativas individuales ($p=0,003$; $V \text{ Cramer}=0,641$). Ya se dio cuenta de la posible explicación de esta asociación en el análisis de las asociaciones entre estas variables para el anterior estrato. Pero en este caso las asociaciones locales entre las diferentes categorías de las variables dependientes e independientes no son las que se producían en ese caso: ahora las negativas se asocian con el tercil jerárquicamente inferior (que agrupa a los terciles inferior y medio) y no se producen en el caso del tercil superior. No se ha encontrado ninguna posible explicación a este fenómeno. En este estrato la última asociación se establece entre el porcentaje de ocupación y los hogares sin personas localizables —asociación recurrente a lo largo de los anteriores 4 estratos analizados—. En esta ocasión la asociación, que es muy significativa e intensa ($p=0,019$; $V \text{ Cramer}=0,533$), también parece adoptar una extraña forma, ya que el tercil jerárquicamente superior (que se trata del medio) no se asocia con esta incidencia, que se produce para las proporciones de ocupación menores (entre el 22,29% y el 37,60% de los ocupados de una sección censal).

Discusión de los resultados

Los análisis efectuados durante el primer nivel de análisis, consagrado al examen de las asociaciones bivariadas establecidas entre variables dependientes y, alternativamente, independientes e intermedias, se han caracterizado por lograr unas significatividades acordes al nivel de confianza fijado a lo largo de nuestra investigación —que era del 90% —, es decir, entre $p=0,110$ y $p=0,007$; y por alcanzar unas intensidades de asociación que oscilaban entre $V \text{ Cramer}=0,100$ y $V \text{ Cramer}=0,203$. Durante el segundo nivel de análisis, centrado en la exploración de las asociaciones que ligaban a las variables intermedias con las independientes, se obtuvieron asociaciones que se concretaron en unas significatividades que iban mucho más allá del nivel de confianza delimitado en nuestro estudio (ya que en este caso las asociaciones eran significativas para un $NC=99,9\%$) y en unas intensidades de relación superiores a las anteriores, en tanto que fluctuaban entre $V \text{ Cramer}=0,348$ y $V \text{ Cramer}=0,425$. A lo largo del tercer nivel de análisis, en que se testaron las asociaciones trivariadas entre las variables dependientes e independientes para cada uno de los 5 tipos de municipio considerados, se consiguieron unas significatividades de las asociaciones que estaban entre $p=0,110$ y $p=0,003$ y unas intensidades de asociación de las más relevantes de entre todas las conseguidas a lo largo de los 3 niveles de análisis, que oscilaban entre $V \text{ Cramer}=0,230$ y $V \text{ Cramer}=0,641$. Es decir, ha sido a lo largo de este tercer nivel de análisis que se han obtenido las significatividades más adecuadas y las intensidades de relación entre variables dependientes e independientes más notables.

La conclusión de nuestros resultados es, pues, la necesidad de llevar a cabo la descripción de las incidencias considerando los dos niveles descriptivos que hemos tomado en cuenta a lo largo del tercer nivel de análisis. Estos dos niveles deben relacionarse necesariamente. El primer nivel viene representado por una característica como es el número de habitantes del municipio; mientras que el segundo nivel se refiere a las características sociodemográficas de la sección censal donde se producen y registran las incidencias. En el balance de los resultados de este análisis descriptivo deben tenerse en cuenta dos polos opuestos. El polo positivo se refiere a la relativa homogeneidad de las asociaciones significativas obtenidas, ya que, en la mayor parte de los casos, una misma variable independiente solía asociarse con una misma variable dependiente, o, en algunos casos, con variables dependientes muy parecidas entre sí. De la misma forma se refiere a la coherencia de la mayoría de las asociaciones globales que los análisis han señalado como significativas. El polo negativo se refiere a la ambigüedad de las asociaciones locales que se han establecido en algunos casos. En este sentido, por estratos, las mayores dificultades para explicar las asociaciones locales obtenidas y también algunas asociaciones globales entre variables independientes y dependientes se han registrado en las «capitales de provincia» y los «otros municipios de más de 50000 habitantes», que eran las categorías que acumulaban el mayor número de asociaciones significativas en el tercer nivel de análisis; los «municipios de 10001 a 50000 habitantes» no han presentado muchos obstáculos; así como tampoco —aunque sí unos pocos más— los «municipios de 2001 a 10000 habitantes». En los «municipios de hasta 2000 habitantes» no se han observado asociaciones de ningún tipo. En general, se ha conseguido hallar explicaciones satisfactorias para las asociaciones globales obtenidas y se han presentado ligeros inconvenientes a nivel de las asociaciones locales entre categorías de variables. Para solventar estos inconvenientes se han propuesto las siguientes medidas.

Es recomendable que, de cara a futuras investigaciones, se considere la necesidad de introducir nuevas variables independientes, aparte de las que han sido consideradas aquí; de estudiar la pertinencia de las consideradas, y, en muchos casos, de revisar la pertinencia de algunas de las asociaciones que se han revelado significativas. Apuntamos también la necesidad de considerar una variable muy importante que no ha sido considerada en nuestro estudio: la

población inactiva, ya sea de forma separada (mayores de 65 años; estudiantes a partir de 16 años; personas menores de 16 años; y otra población inactiva de entre 16 y 64 años, cada grupo aisladamente) o agrupada. Se trata de una importante variable ya que, en nuestro caso, marca el punto de inflexión entre la población encuestable y la no encuestable, y este hecho puede tener consecuencias muy claras en relación a alguna de las 7 incidencias que han sido consideradas por nosotros.

También es deseable que se estudien las incidencias desde el punto de vista de las relaciones que se establecen entre las variables independientes. Estas relaciones no han sido consideradas en nuestra investigación y, como se ha visto a través de diversos ejemplos (caso de las asociaciones, detectadas en el primer nivel de análisis, entre las imposibilidades por idioma y la ocupación como variable independiente; caso de las asociaciones, observadas durante el tercer nivel de análisis en los otros municipios de más de 50000 habitantes, entre la ausencia de persona seleccionada en el hogar y la proporción de inmigrantes), parecería que en muchos casos pueden explicar —a través de relaciones espurias entre variables— las asociaciones, en muchos casos inexplicables, que hemos señalado en algunas ocasiones.

El anterior punto podría generalizarse en la necesidad de estudiar las relaciones que se establecen entre variables dependientes (incidencias) e independientes a la luz de las dinámicas sociales (en cada tipo de municipio, ya que estas varían en cada caso) que ayudan a explicar estas relaciones, ya que en muchos casos —especialmente durante el tercer nivel analítico: recuérdense los casos de la asociación entre la ocupación y la ausencia de la persona seleccionada o entre la tasa de envejecimiento de la población de una sección censal y la población no encuestable— las asociaciones locales entre categorías de variables dependientes e independientes resultan inexplicables si no se profundiza en las causas de estas asociaciones. Para estudiar estas relaciones deberían considerarse muy especialmente las relaciones que se establecen entre las variables y la variable central que marca la inclusión o el rechazo de población: en nuestro caso la población diana era la población ocupada; y, por tanto, deberían estudiarse muy especialmente las relaciones entre la ocupación y las demás variables independientes consideradas (muy especialmente la población inactiva). La principal variable independiente que, como se mostró a lo largo del tercer nivel de análisis especialmente (aunque no sólo en éste), parecía presentar una relación poco clara, que debería estudiarse apropiadamente, con la proporción de ocupados era la proporción de inmigrantes en el caso de incidencias como la ausencia de persona seleccionada en hogar.

Agotados los anteriores frentes en relación a determinadas incongruencias —al menos, aparentes— de los resultados obtenidos, y agotada también la posibilidad de cuestionar la estrategia muestral —que ha sido expuesta con bastante detalle en el primer apartado de este artículo, y que parece la adecuada—, en último lugar cabría preguntarse sobre la fiabilidad de los resultados obtenidos. Es decir, ¿es suficiente una muestra de 254 secciones censales para estudiar las incidencias que se producen durante el trabajo de campo con encuestas personales? ¿O un número tan reducido de elementos puede acabar por arrojar resultados contradictorios? Y, por tanto, ¿deberíamos poder contar con mayores tamaños muestrales? Estas y las otras preguntas que hemos puesto sobre la mesa deberían ser contestadas a la luz de nuevas investigaciones cada vez con un mayor bagaje analítico a nuestras espaldas.

Referencias bibliográficas

- Cea D'Ancona, M. A. (1999). *Metodología cuantitativa. Estrategias y técnicas de investigación social*. Madrid: Editorial Síntesis.
- Generalitat de Catalunya. (2005). “Identificació i avaluació de riscos psicosocials. Manual de la versió mitjana del mètode PSQ CAT21 COPSOQ (Qüestionari psicosocial de Copenhaguen) adaptat”. In: *Manual per a la identificació i avaluació de riscos laborals*. Barcelona: Departament de Treball i Indústria. Páginas 155-202. En formato electrónico: http://www.gencat.net/treballiindustria/relacions_laborals/seguretatisalut/manual_riscos/psqcat21/index.html.
- Moncada, S.; Llorens, C.; Navarro, A.; Kristensen, T. S. (2005). ISTAS21: Versión en lengua castellana del cuestionario psicosocial de Copenhague (COPSOQ). *Arch Prev Riesgos Labor*, nº 8, vol. 1, 18-29.
- Moncada, S.; Llorens, C.; Kristensen, T. S.; Vega, S. (2006). El método COPSOQ (ISTAS21, PSQCAT21) de evaluación de riesgos psicosociales. *Nota Técnica de Prevención*, nº 703. Madrid: Instituto Nacional de Seguridad e Higiene en el Trabajo.

Anexos

ANEXO 1: Variables consideradas en los análisis efectuados

Tabla 1: Resumen de las variables

Tipo de variable	Variable	Casos válidos (secciones censales)	Casos perdidos (secciones censales)	Casos totales (secciones censales)
«Intermedias»	Tipo de municipio (capital de provincia / no capital)	254	0	254
	Categorización por estratos (número de habitantes en municipio) (5 estratos)	254	0	254
	Categorización por estratos (2 estratos)	254	0	254
«Independientes» (*)	% mayores de 65 años en sección censal (en terciles)	242	12	254
	% inmigrantes en sección censal (en terciles)	242	12	254
	% ocupados de 16 años o más en sección censal (en terciles)	242	12	254
«Dependientes» (*)	No contactos de hogar (sin incidencias / con incidencias)	254	0	254
	Negativas individuales (sin incidencias / con incidencias)	254	0	254
	Hogares sin personas localizables (sin incidencias / con incidencias)	254	0	254
	Ausencias de persona seleccionada en hogar (sin incidencias / con incidenc.)	254	0	254
	Imposibilidad por idioma (sin / con)	254	0	254
	Otro tipo de imposibilidad (sin / con)	254	0	254
	Población no encuestable (sin / con)	254	0	254

* A pesar de que no sea propio postular relaciones de dependencia e independencia en el contexto del análisis de tablas de contingencia, se ha usado esta denominación.

Tabla 2: Correlaciones entre variables intermedias e independientes

		Categor. por estratos (número de habitantes en municipio) (5 estratos)	Tipo de municipio	Categor. por estratos (2 estratos)	% mayores de 65 años en sección censal	% inmigra. en sección censal	% ocupados de 16 años o más en sección censal
Categor. por estratos (número de habitantes en municipio) (5 estratos)	Correlación de Pearson	1	-,791(**)	-,560(**)	,089	,215(**)	-,078
	Sig. (bilateral)		,000	,000	,167	,001	,224
Tipo de municipio	Correlación de Pearson	-,791(**)	1	,217(**)	,172(**)	-,159(*)	,000
	Sig. (bilateral)	,000		,000	,007	,013	1,000
Categor. por estratos (2 estratos)	Correlación de Pearson	-,560(**)	,217(**)	1	-,269(**)	-,134(*)	,135(*)
	Sig. (bilateral)	,000	,000		,000	,038	,036
% mayores de 65 años en sección censal	Correlación de Pearson	,089	,172(**)	-,269(**)	1	-,081	-,384(**)
	Sig. (bilateral)	,167	,007	,000		,211	,000
% inmigrantes en sección censal	Correlación de Pearson	,215(**)	-,159(*)	-,134(*)	-,081	1	,433(**)
	Sig. (bilateral)	,001	,013	,038	,211		,000
% ocupados de 16 años o más en sección censal	Correlación de Pearson	-,078	,000	,135(*)	-,384(**)	,433(**)	1
	Sig. (bilateral)	,224	1,000	,036	,000	,000	
	Sig. (bilateral)	,000	,000	,006	,054	,003	1,000

** La correlación es significativa al nivel 0,01 (bilateral); * La correlación es significativa al nivel 0,05 (bilateral).

ANEXO 2: Tablas de asociaciones significativas entre variables, detectadas en el análisis de tablas de contingencia bivariado

Tabla 3: Asociaciones globales

Variable independiente	Variable dependiente	χ^2	Significación asintótica (bilateral)	V Cramer
Tipo de municipio	Negativas individuales	2,55 8	0,110	0,100 (0)
Tipo de municipio	Ausencias de persona seleccionada en hogar	5,78 8	0,016	0,151 (1)
Categorización por estratos (2 estratos)	No contactos de hogar	6,18 5	0,013	0,156 (2)
Categorización por estratos (2 estratos)	Negativas individuales	5,12 5	0,024	0,142 (3)
Categorización por estratos (2 estratos)	Otro tipo de imposibilidad	4,44 8	0,035	0,132 (4)
% Ocupados > = 16 años	Ausencias de persona seleccionada	9,93 7	0,007	0,203 (5)
% Ocupados > = 16 años	Imposibilidades por idioma	6,65 9	0,036	0,166 (6)
% Mayores de 65 años	Otro tipo de imposibilidad	4,49 5	0,106	0,136 (7)
% Mayores de 65 años	Población no encuestable	4,96 3	0,084	0,143 (8)
% Ocupados > = 16 años	Población no encuestable	4,89 7	0,086	0,142 (9)

Tabla 3.1: Asociaciones locales entre categorías de variables

Var. ind: variable y categoría	Var. dep: variable y categoría	Residuos tipificados
Tipo de municipio: No capital de provincia	Negativas individuales: Sin incidencias	1,6 (0) (*)
Tipo de municipio: Capital de provincia	Negativas individuales: Con incidencias	1,6 (0)
Tipo de municipio: No capital de provincia	Ausencias de persona seleccionada en hogar: Sin incid.	2,4 (1)
Tipo de municipio: Capital de provincia	Ausencias de persona seleccionada en hogar: Con incid.	2,4 (1)
Categorización por estratos (2): Munic. de hasta 2000 h.	No contactos de hogar: Sin incidencias	2,5 (2)
Categorización por estratos (2): Munic. de 2001 y más h.	No contactos de hogar: Con incidencias	2,5 (2)
Categorización por estratos (2): Munic. de hasta 2000 h.	Negativas individuales: Sin incidencias	2,3 (3)
Categorización por estratos (2): Munic. de 2001 y más h.	Negativas individuales: Con incidencias	2,3 (3)
Categorización por estratos (2): Munic. de hasta 2000 h.	Otro tipo de imposibilidad: Sin incidencias	2,1 (4)
Categorización por estratos (2): Munic. de 2001 y más h.	Otro tipo de imposibilidad: Con incidencias	2,1 (4)
% Ocupados > = 16 años: Del 22,29% al 37,60%	Ausencias de persona seleccionada: Sin incidencias	1,9 (5)
% Ocupados > = 16 años: Del 43,53% al 62,50%	Ausencias de persona seleccionada: Sin incidencias	1,2 (5) (pa=77%) (**)
% Ocupados > = 16 años: Del 37,61% al 43,52%	Ausencias de persona seleccionada: Con incidencias	3,1 (5)
% Ocupados > = 16 años: Del 22,29% al 43,52% (3*)	Imposibilidades por idioma: Sin incidencias	2,5 (6)
% Ocupados > = 16 años: Del 43,53% al 62,50%	Imposibilidades por idioma: Con incidencias	2,6 (6)
% Mayores de 65 años: Del 13,14% al 47,97% (4*)	Otro tipo de imposibilidad: Sin incidencias	2,2 (7)
% Mayores de 65 años: Del 2,60% al 13,13%	Otro tipo de imposibilidad: Con incidencias	2,1 (7)
% Mayores de 65 años: Del 19,87% al 47,97%	Población no encuestable: Sin incidencias	1,3 (8) (pa=80%) (**)
% Mayores de 65 años: Del 13,14% al 19,86%	Población no encuestable: Con incidencias	2,2 (8)
% Ocupados > = 16 años: Del 37,61% al 43,52%	Población no encuestable: Sin incidencias	2,0 (9)
% Ocupados > = 16 años: Del 43,53% al 62,50%	Población no encuestable: Con incidencias	2,0 (9)

* Los números entre paréntesis —(0) en este caso— de las tablas de asociaciones locales, en todas las tablas de los anexos, identifican las asociaciones locales entre categorías de variables que corresponden a cada una de las asociaciones globales que se señalan, con el mismo número, en las tablas de asociaciones globales.

** pa= Probabilidad de asociación. Estas asociaciones no son significativas para un nivel de significación del 0,10 (nivel de confianza=90%).

3* Se fundió el primer tercil (del 22,29% de ocupados al 37,60%) con el segundo (del 37,61% de ocupados al 43,52%).

4* Se fundió el segundo tercil (del 13,14% de mayores de 65 años al 19,86%) con el último (del 19,87% de mayores de 65 años al 47,97%).

ANEXO 3: Tablas de asociaciones significativas entre variables intermedias (caracterización del municipio según el número de habitantes) y variables independientes (caracterización sociodemográfica de las secciones censales), detectadas en el análisis de tablas de contingencia bivariado

Tabla 4: Asociaciones globales entre variables independientes e intermedias (*)

Variable intermedia	Variables independientes	χ^2	Significación asintótica (bilateral)	V Cramer
Categorización por estratos (5 estr.)	% mayores de 65 años	43,66 6	0,000	0,425 (10)
Categorización por estratos (5 estr.)	% inmigrantes	29,32 8	0,000	0,348 (11)

* En estos análisis no se consideraron las variables: Tipo de municipio, por presentar un alto coeficiente de correlación de Pearson (-0,791, significativo al nivel 0,01) con la variable Categorización por estratos (5); y Categorización por estratos (2) por tratarse de una categorización distinta de la misma variable.

Tabla 4.1: Asociaciones locales entre variables intermedias e independientes

Var. intermedia: variable y categoría	Var. independiente: variable y categoría	Residuos tipificados
Categorización por estratos (5): Capitales de provincia	% mayores de 65 años: Del 19,87% al 47,97%	1,7 (10) (pa=91%)
Categorización por estratos (5): Otros municip. > 50000 h.	% mayores de 65 años: Del 2,60% al 13,13%	3,6 (10)
Categorización por estratos (5): Munic.10001 a 50000 h.	% mayores de 65 años: Del 2,60% al 13,13%	2,7 (10)
Categorización por estratos (5): Munic. 2001 a 10000 h.	% mayores de 65 años: Del 13,13% al 47,97% (*)	1,8 (10) (pa=93%)
Categorización por estratos (5): Munic. de hasta 2000 h.	% mayores de 65 años: Del 19,87% al 47,97%	4,1 (10)
Categorización por estratos (5): Capitales de provincia	% inmigrantes: Del 6,56% al 10,80%	3,4 (11)
Categorización por estratos (5): Munic.10001 a 50000 h.	% inmigrantes: Del 10,81% al 48,91%	2,3 (11)
Categorización por estratos (5): Munic. 2001 a 10000 h.	% inmigrantes: Del 10,81% al 48,91%	2,0 (11)
Categorización por estratos (5): Munic. de hasta 2000 h.	% inmigrantes: Del 10,81% al 48,91%	1,5 (11) (pa=87%) (**)

* Se fundió el segundo tercil (del 13,14% de mayores de 65 años al 19,86%) con el último (del 19,87% de mayores de 65 años al 47,97%).

** pa= Probabilidad de asociación. Estas asociaciones no son significativas para un nivel de significación del 0,10 (nivel de confianza=90%).

ANEXO 4: Tablas de asociaciones significativas entre variables, detectadas en el análisis de tablas de contingencia trivariado (1 variable dependiente y 1 variable independiente según cada categoría de la variable intermedia considerada «Categorización por estratos (5)»)

Tabla 5: Asociaciones globales

Categoría de variable intermedia	Variable independiente	Variable dependiente	χ^2	Significación asintótica (bilateral)	V Cramer
Capitales de provincia	% inmigrantes	Hogares sin personas localizables	5,764	0,056	0,250 (12)
	% ocupados	Hogares sin personas localizables	5,802	0,055	0,251 (13)
	% ocupados	Ausencias de persona seleccionada en hogar	9,173	0,010	0,316 (14)
	% > de 65 años	Población no encuestable	4,850	0,088	0,230 (15)
Otros municipios de más de 50000 habitantes	% inmigrantes	No contactos	4,503	0,105	0,327 (16)
	% ocupados	Hogares sin personas localizables	4,834	0,089	0,339 (17)
	% inmigrantes	Ausencias de persona seleccionada en hogar	8,085	0,018	0,439 (18)
	% ocupados	Ausencias de persona seleccionada en hogar	8,273	0,016	0,444 (19)
Municipios de 10001 a 50000 habitantes	% inmigrantes	No contactos	5,079	0,079	0,286 (20)
	% ocupados	Negativas individuales	5,159	0,076	0,288 (21)
	% ocupados	Hogares sin personas localizables	6,326	0,042	0,319 (22)
	% ocupados	Población no encuestable	4,408	0,110	0,267 (23)
Municipios de 2001 a 10000 habitantes	% inmigrantes	No contactos	7,570	0,023	0,520 (24)
	% > de 65 años	Negativas individuales	6,965	0,031	0,499 (25)
	% ocupados	Negativas individuales	11,516	0,003	0,641 (26)
	% ocupados	Hogares sin personas localizables	7,955	0,019	0,533 (27)

Tabla 5.1: Asociaciones locales

Categoría de variable intermedia	Variable independiente y categoría	Variable dependiente y categoría	Residuos tipificados
Capitales de provincia	% inmigrantes: Del 0,51% al 6,55%	Hogares sin personas localizables: Sin incid.	2,1 (12)
	% inmigrantes: 6,56% - 48,91% (3*)	Hogares sin personas localizables: Con incid.	2,5 (12)
	% ocupados: Del 22,29% - 43,52% (2*)	Hogares sin personas localizables: Sin incid.	2,2 (13)
	% ocupados: Del 43,53% al 62,50%	Hogares sin personas localizables: Con incid.	2,2 (13)
	% ocupados: Del 37,61% al 43,52%	Ausencia persona seleccionada: Con incid.	3,0 (14)
	% ocupados: Del 43,53% al 62,50%	Ausencia persona seleccionada: Sin incid.	2,0 (14)
	% mayores 65: Del 13,14% al 19,86%	Población no encuestable: Con incidencias	2,2 (15)
	% mayores 65: Del 19,87% al 47,97%	Población no encuestable: Sin incidencias	1,6 (15)
Otros municipios de más de 50000 habitantes	% inmigrantes: Del 6,56% al 10,80%	No contactos: Sin incidencias	1,8 (16)
	% inmigrantes: Del 0,51% al 6,55%	No contactos: Con incidencias	1,9 (16)
	% ocupados: Del 22,29% - 43,52% (2*)	Hogares sin personas localizables: Con incid.	2,4 (17)
	% ocupados: Del 43,52% al 62,50%	Hogares sin personas localizables: Sin incid.	2,1 (17)
	% inmigrantes: Del 0,51% al 6,55%	Ausencia persona seleccionada: Sin incid.	2,8 (18)
	% inmigrantes: Del 6,56 - 48,91% (3*)	Ausencia persona seleccionada: Con incid.	2,9 (18)
	% ocupados: Del 22,29% al 37,60%	Ausencia persona seleccionada: Sin incid.	2,8 (19)
	% ocupados: Del 37,61 - 62,50% (4*)	Ausencia persona seleccionada: Con incid.	2,6 (19)

Municipios de 10001 a 50000 habitantes	% inmigrantes: Del 6,56% al 10,80%	No contactos: Sin incidencias	2,2 (20)
	% inmigrantes: Del 10,81% al 48,91%	No contactos: Con incidencias	1,1 (20)
	% ocupados: Del 37,61% al 43,52%	Negativas indiv.: Sin incidencias	2,3 (21)
	% ocupados: Del 43,53% al 62,50%	Negativas indiv.: Con incidencias	1,4 (21)
	% ocupados: Del 22,29% al 37,60%	Hogares sin personas localizables: Sin incid.	2,4 (22)
	% ocupados: Del 37,61% - 62,50% (4*)	Hogares sin personas localizables: Con incid.	2,4 (22)
	% ocupados: Del 22,29% - 43,52% (2*)	Población no encuestable: Sin incidencias	2,2 (23)
	% ocupados: Del 43,53% al 62,50%	Población no encuestable: Con incidencias	2,1 (23)
Municipios de 2001 a 10000 habitantes	% inmigrantes: Del 6,56% al 10,80%	No contactos: Con incidencias	2,7 (24)
	% inmigrantes: Del 10,81% al 48,91%	No contactos: Sin incidencias	1,6 (24)
	% mayores 65: Del 2,60% - 19,86% (*)	Negativas indiv.: Sin incidencias	2,9 (25)
	% mayores 65: Del 19,87% - 47,97%	Negativas indiv.: Con incidencias	2,6 (25)
	% ocupados: Del 22,49% - 43,52% (2*)	Negativas indiv.: Con incidencias	3,0 (26)
	% ocupados: Del 43,53% al 62,50%	Negativas indiv.: Sin incidencias	3,1 (26)
	% ocupados: Del 22,29% al 37,60%	Hogares sin personas localizables: Con incid.	2,1 (27)
	% ocupados: Del 37,61% al 43,52%	Hogares sin personas localizables: Sin incid.	2,7 (27)

* Se fundió el primer tercil (del 2,60% de mayores de 65 años al 13,13%) con el segundo (del 13,13% de mayores de 65 años al 19,86%).

2* Se fundió el primer tercil (del 22,29% de ocupados al 37,60%) con el segundo (del 37,61% de ocupados al 43,52%).

3* Se fundió el segundo tercil (del 6,56% de inmigrantes al 10,80%) con el tercero (del 10,81% de inmigrantes al 48,91%).

4* Se fundió el segundo tercil (del 37,61% de ocupados al 43,52%) con el tercero (del 43,53% de ocupados al 62,50%).