

# *Hacia la recuperación del contexto en epidemiología: variables y falacias en el análisis multinivel<sup>1</sup>*

Ana V. Diez Roux

**A** LO LARGO DE LA HISTORIA DE LA SALUD PÚBLICA, según la teoría causal de la enfermedad más aceptada en cada época, se consideraron distintos aspectos de los individuos o de su ambiente como posibles "causas" de enfermedad (1-3). En sus orígenes, la salud pública era esencialmente ecológica y relacionaba la salud y la enfermedad con el medio ambiente y las características de la comunidad (4-8). Posteriormente, con el advenimiento de la teoría de los gérmenes y la teoría etiológica asociada de unicausalidad de la enfermedad, los organismos infecciosos se tornaron el factor "ambiental" relevante (9). Otros aspectos eran importantes en la medida que llevaban a que se reprodujeran o transmitieran las "causas" biológicas de enfermedad (10). En el siglo XX el aumento de importancia de las enfermedades crónicas hizo que se buscaran nuevos factores causales. El interés se desplazó de los factores ambientales a los factores de nivel individual y la investigación se enfocó hacia las características comportamentales y biológicas como factores de riesgo de las enfermedades crónicas.

Las causas de enfermedad, que se buscaban antes en el ambiente en su conjunto, pasaron a buscarse en factores específicos del ambiente (organismos biológicos) y conductas de los individuos. Partiendo de un modelo de determinación relativamente vago, holístico, el modelo causal se desplazó hacia la unicausalidad de la teoría de los gérmenes y hacia la multicausalidad (la "red de causación") prevalente en la actualidad, en la que se supone que diversos factores de riesgo, tanto biológicos como conductuales, interactúan en la causación de la enfermedad (11). Este proceso de cambio del modelo de causación de la enfermedad ha ido acompañado de una "individualización" progresiva del riesgo.

Esta individualización del riesgo perpetúa la idea de que el riesgo se determina individual y no socialmente y hace que se mire con desinterés la investigación del efecto de variables de nivel macro o variables grupales en los resultados correspondiente a individuos. Los "estilos de vida" y las "conductas personales" se consideran opciones libres de los individuos, disociadas de los contextos sociales que los delimitan y restringen (12). Esta tendencia que explica los patrones de salud y enfermedad exclusivamente por las características de los individuos es análoga a la doctrina del individualismo metodológico en ciencias sociales. Según esta doctrina, "los hechos referentes a la sociedad y los fenómenos sociales han de explicarse exclusivamente por las características de los individuos" (13, p. 77). El correlato lógico es que cualquier variable se mide mejor a nivel individual que a nivel macro o grupal, porque se supone que el individuo es el

---

<sup>1</sup> Publicado originalmente en inglés en *American Journal of Public Health*, Vol. 88, No. 21, 1998, pp. 216-222.

factor verdaderamente importante en la causación de la enfermedad. Las variables grupales se incluyen en el análisis solamente como aproximaciones burdas a los datos de nivel individual cuando estos no están disponibles. Como se discutirá en lo que sigue, ignorar el papel de las variables de nivel grupal o macro puede llevar a una comprensión incompleta de los determinantes de enfermedad en los individuos, así como en las poblaciones. Las variables macro o grupales afectan a los individuos directamente y también restringen sus opciones.

Hay varias formas de oponerse al individualismo metodológico prevalente en la investigación epidemiológica actual. Por una parte, la interpretación de los efectos a nivel individual debe tener presente la relación con los procesos a nivel macro. Muchas variables medidas a nivel individual están fuertemente condicionadas por procesos sociales que operan a nivel de grupo social o de toda la sociedad. Otra posibilidad es incluir variables de niveles macro y micro en los estudios epidemiológicos, como han hecho ya diversos autores (12,14-17). La inclusión de variables de nivel grupal o macro (junto con variables de nivel individual) en la investigación sanitaria plantea problemas difíciles tanto desde el punto de vista metodológico como desde el punto de vista teórico. Para que sea significativa (y no sólo una simple adición de un conjunto de variables a la "red causal") requiere el desarrollo de modelos causales de enfermedad (e hipótesis verificables) que abarquen diversos niveles y expliquen cómo las variables de nivel individual y grupal conforman la salud y la enfermedad conjuntamente.

En el resto de este trabajo se discutirán más en detalle algunos conceptos básicos y problemas involucrados en los análisis multinivel.

## *Análisis contextual o multinivel*

Muchos sociólogos han argüido que en las vidas de los individuos no sólo influyen sus características personales sino también las características de los grupos sociales a los que los individuos pertenecen. La idea que subyace es que los grupos sociales son unidades legítimas de análisis, que las propiedades grupales son distintas a las propiedades de los individuos y que estas macrovariables pueden actuar sobre el efecto general independientemente de las características individuales o modificar el efecto de las características individuales (18, 19). Por consiguiente, para entender las conductas y los resultados individuales puede ser útil analizar no sólo las características de los individuos sino también las de los grupos sociales a los cuales pertenecen (15).

El estudio de los efectos de las características colectivas o grupales en los resultados a nivel individual ha sido denominado análisis contextual (18, 20-22) o análisis "de niveles múltiples", o multinivel<sup>2</sup> (14, 23, 24). Los efectos de las propiedades de nivel grupal en resultados de nivel individual se han denominado efectos contextuales. Empíricamente, el análisis contextual implica la incorporación de variables grupales en ecuaciones de nivel micro apropiadas para estudiar personas como unidades de análisis (18) (Tabla 1). Susser ha llamado estos estudios "estudios mixtos" porque en ellos se investigan los efectos de variables independientes

---

<sup>2</sup> En inglés, *contextual analysis* y *multi level analysis*, respectivamente (N. del T.)

ecológicas sobre variables dependientes de nivel individual (15). En estos análisis se han usado diversos términos como sinónimos para designar las variables de nivel grupal, por ejemplo variables ecológicas (15, 25), variables de nivel macro<sup>3</sup> (14, 21), variables contextuales (18, 19, 21, 26), y variables agregadas (14). Como los términos "variable contextual" y "variable agregada" se han aplicado específicamente a un tipo especial de variable de nivel grupal (14, 15, 25) no se usarán en este trabajo. Las expresiones "variable a (o de) nivel grupal", "variable a (o de) nivel macro", "variable de macronivel" y "variable ecológica" se usarán como sinónimos.

Los sociólogos han documentado los efectos del nivel grupal en diversos efectos a nivel individual, por ejemplo el comportamiento electoral (27), el rendimiento educativo (28), las actitudes hacia la educación (29) y la satisfacción vital (30). En epidemiología es menos habitual que se use y se discuta el análisis multinivel, pero diversos autores han hecho hincapié en la importancia de incluir variables de nivel macro junto a variables de nivel individual en los estudios epidemiológicos. Según Morgenstern, "aunque el desarrollo de la enfermedad es un fenómeno biológico individual, es posible que ciertos determinantes importantes de la enfermedad no se puedan operacionalizar completamente a nivel individual" (31, p. 16). Von Korff *et al* (14), han sugerido que se incluyan múltiples niveles de determinación (variables micro y macro) en los estudios epidemiológicos de enfermedades crónicas. Otros autores han enfatizado la necesidad de investigar cómo las medidas cuantitativas ecológicas del contexto socioeconómico conforman la distribución de los factores de riesgo (32,33) y articulan "las conexiones entre las acciones de los individuos y el contexto socioecológico en que esas acciones se realizan (12, p. 818). En recientes revisiones del uso de los estudios ecológicos en epidemiología, Susser (15) y Schwartz (34) han sugerido que las variables ecológicas pueden dar información que no captan los datos de nivel individual. Otros autores han defendido un nuevo paradigma epidemiológico que enfoca las relaciones internas y externas de estructuras organizadas en una jerarquía de niveles y que parte de la necesidad de incorporar esos múltiples niveles al análisis epidemiológico (11, 35).

La idea de que los factores de nivel grupal pueden ser importantes en la determinación de resultados de nivel individual no es nueva inclusive para la epidemiología. Ya en 1916, en un ejemplo clásico de investigación epidemiológica, Goldberger *et al.* (36) descubrieron que el riesgo de pelagra se relacionaba con el acceso general de la villa o aldea a frutas y verduras frescas, independientemente del nivel de ingreso de los individuos. Además, en los núcleos de población donde la disponibilidad de alimentos frescos no era adecuada, los hogares de ingreso elevado parecían estar menos protegidos de la pelagra que allá donde la disponibilidad de alimentos era adecuada, lo que ilustra la interacción entre factores a nivel individual y a nivel de toda la población. Otro ejemplo clásico puede encontrarse en el campo de las enfermedades infecciosas, en el que se acepta desde hace mucho que el riesgo individual de contraer una infección depende no sólo del estado inmunológico del individuo sino también del estado inmunitario de la comunidad.

---

<sup>3</sup> En inglés se habla de *macro-level variable* o, simplemente, *macro variable*. Además de las traducciones literales "Variable de nivel macro" o "variable macro", también parecen posibles traducciones como "Variable de macronivel" o, incluso, "macrovariable" (N. del T.).

En los últimos años hubo un aumento considerable de las publicaciones que resaltan la importancia de las variables grupales para entender las desigualdades sociales en salud (16, 37, 38). Se descubrió que diversas medidas zonales concretas de pobreza o carencias, una vez considerado el efecto del nivel individual, influyen sobre los resultados de salud (39-46). Otros estudios han sugerido que las variables de nivel de vecindario también pueden conformar la distribución de conductas relacionadas con la salud (12, 47-49), aunque en otros estudios se encontraron pocos datos a favor de esa relación (50,51).

El análisis multinivel también se ha usado en otros campos de la salud pública. Así se ha demostrado que las características generales del vecindario están relacionadas con la violencia doméstica (52) y los demógrafos han mostrado cómo las variables que corresponden al país en conjunto, por ejemplo el producto nacional bruto per cápita, se relacionan con los índices de fertilidad independientemente de los factores de nivel individual. (53). Se descubrió que las variables a nivel de comunidad modifican los efectos de la educación materna en el riesgo de diarrea infantil (54). El análisis multinivel se ha usado para estudiar el efecto del tipo de servicio de salud en la inmunización infantil (56) y Koopman *et al* (55). ilustraron la interacción entre variables grupales e individuales en la transmisión del dengue.

### *Tipos de variables de nivel grupal utilizadas en el análisis contextual o multinivel*

La razón para incorporar variables grupales en el análisis contextual o multinivel es que estas variables proporcionan información que no captan los datos individuales. Por ejemplo, el ingreso promedio de un vecindario puede proporcionar información que no capta el ingreso a nivel individual. El ingreso promedio puede ser indicativo de factores del vecindario potencialmente relacionados con la salud (como la existencia de espacios recreativos, la calidad escolar, las condiciones de las calles, las condiciones del medio ambiente, los tipos de alimentos disponibles, etc.). Estos factores pueden afectar a todos los miembros de la comunidad, sea cual sea su ingreso individual. De forma semejante, el nivel de desempleo de la comunidad puede afectar a todos los individuos de una comunidad, independientemente de que estén o no desempleados.

La distinción conceptual entre las variables del nivel grupal e individual es importante cuando: 1) hay variables análogas en ambos niveles, pero cada una mide un constructo diferente (como en los ejemplos citados); 2) la variable sólo puede medirse a nivel grupal, porque sólo está definida a nivel grupal (caso por ejemplo de la existencia de ciertas regulaciones, o medidas como la densidad poblacional, el grado de desigualdad de ingreso en una comunidad, el régimen político, el *status* legal de las mujeres, etc.).

Las variables que reflejan características grupales han sido clasificadas en dos tipos básicos (14,19,21, 25,26, 57): variables derivadas y variables integrales. Las variables derivadas (también llamadas variables analíticas o agregadas) resumen las características de los individuos del grupo (como promedios, proporciones o medidas de dispersión; por ejemplo, el porcentaje de personas con escuela secundaria incompleta, el ingreso medio de los hogares, la desviación

estándar de la distribución del ingreso). Susser se ha referido a variables derivadas denominándolas variables contextuales o de contexto (15). Las variables integrales (también llamadas variables primarias o globales) describen características del grupo que no derivan de las características de sus miembros (por ejemplo, que existan o no ciertas regulaciones, que se disponga de atención sanitaria, que exista un tipo determinado de sistema político, o la densidad de población). Las variables integrales no tienen análogo a nivel individual. Un subconjunto especial de las variables derivadas es el promedio de la variable dependiente en el grupo. Al efecto contextual sobre un resultado de varias variables dependientes agrupadas, Susser lo llama "contagio" (15). Este uso especial de las variables derivadas es particularmente pertinente en las enfermedades infecciosas, pero también puede ser importante en el estudio de otros resultados de salud o enfermedad. Por ejemplo, la tasa de prevalencia de enfermedad en un grupo modifica la probabilidad de que un determinado individuo adquiera la infección y puede modificar el efecto de las variables a nivel individual en el riesgo de infección (15, 58, 59). Igualmente, la probabilidad de adoptar una cierta conducta ( $Y_{jj}$ ) a menudo depende en parte del grado en que la conducta ya ha sido adoptada en la comunidad ( $Y_i$ ) (21).

Aunque las variables derivadas e integrales a veces se presentan como si fueran conceptualmente distintas, realmente están muy relacionadas. Las variables derivadas en general operan dando forma a ciertas propiedades integrales del grupo. Por ejemplo, la composición de un grupo puede influir en el predominio de ciertos tipos de contactos interpersonales, valores y normas, o puede determinar la existencia de organizaciones o reglamentos del grupo que afecten a todos sus miembros (57). Esencialmente, siempre que las variables derivadas se utilizan en el análisis con textual se supone que captan propiedades del grupo que no son sólo una síntesis de propiedades individuales.

No hay ninguna correspondencia directa entre las denominadas "variables ambientales" (término que se utiliza para referirse a muchos tipos diferentes de variables) y las variables de nivel grupal. El "ambiente social", por ejemplo, puede expresarse tanto a nivel grupal como individual. Un investigador que estudie cómo se relaciona la "participación comunitaria" (una característica del ambiente social) con los resultados sanitarios podría calcular: 1) el número de organizaciones comunitarias a las que cada individuo pertenece (una variable a nivel individual); 2) el porcentaje de personas de la comunidad que pertenece al menos a una organización (variable derivada a nivel grupal); 3) el número de organizaciones comunitarias que existen en la comunidad (variable integral grupal). Cada variable mide un aspecto ligeramente distinto del concepto de participación comunitaria. Otras variables ambientales, por ejemplo la disponibilidad de espacios públicos recreativos, son variables integrales de nivel grupal.

Es más complejo categorizar otras medidas del ambiente físico o químico a nivel grupal (por ejemplo, la contaminación atmosférica o las horas de luz solar en una zona). Se ha dicho que las medidas de exposición física o química a nivel grupal deben considerarse como una categoría separada, distinta tanto de las variables derivadas como de las variables integrales (25). Estas variables no son derivados agregados de características de individuos, pero tienen análogos a nivel grupal e individual (por ejemplo, los días de luz solar a nivel de comunidad y la exposición individual a la luz solar) (25). Sin embargo, las medidas de exposición física o química a nivel grupal en general se utilizan como aproximaciones o sucedáneos de la exposición individual (que

puede ser más difícil de medir por razones logísticas o metodológicas), más que como indicadores de una propiedad grupal conceptualmente diferente de la medida análoga a nivel individual. En ese sentido difieren de las otras variables grupales (derivadas e integrales) que se discuten en este trabajo y que se supone proporcionan información sobre propiedades conceptuales de nivel grupal. Por supuesto que las exposiciones físicas o químicas (medidas a nivel individual o a nivel grupal) pueden tener determinantes a nivel grupal como por ejemplo la presencia de un vertedero de desechos tóxicos (una variable integral a nivel grupal), o el ingreso promedio o la composición racial/étnica del vecindario (una variable derivada de nivel grupal). En una zona los niveles atmosféricos promedio de cierto contaminante pueden depender en gran medida de la presencia de una instalación industrial en esa zona. En este caso, el nivel promedio de contaminación del barrio puede considerarse un indicador de una variable integral grupal (la presencia de esa instalación en el barrio).

### *Problemas planteados por los niveles de análisis: falacias en los estudios que involucran múltiples niveles*

La mayor parte de la literatura referida al problema del nivel de análisis en epidemiología se refiere a la falacias de inferir relaciones en un nivel a partir de datos de otro nivel. Esas falacias surgen cuando "los métodos no se ajustan al modelo conceptual", es decir, cuando el modelo conceptual que se somete a prueba empírica corresponde a un nivel pero los datos corresponden a otro nivel (60). Como los epidemiólogos se dedican (al menos en época reciente) a inferir relaciones a nivel individual, la falacia ecológica o falacia de agregación ha recibido mucha más atención que su acompañante, la falacia atomista. Ambas pueden considerarse variantes del problema que implica hacer inferencias a un nivel de agregación a partir de datos que corresponden a otro nivel. En la falacia ecológica se infiere a nivel individual a partir de datos de nivel grupal. Supongamos por ejemplo que un investigador descubre que el aumento del ingreso per cápita se asocia a nivel del país con un aumento de la mortalidad debida al tráfico de automóviles. Si se infiere que dentro del país el aumento de ingreso se asocia con una mayor mortalidad por tráfico, se puede estar cometiendo la falacia ecológica, porque puede que dentro de cada país, en personas de bajo ingreso la mortalidad debida al tráfico sea siempre mayor que en las personas con ingresos altos. La falacia atomística es la falacia de diseñar inferencias a nivel grupal basadas en datos de nivel individual. En algunos casos lo interesante pueden ser las asociaciones a nivel grupal (no las asociaciones a nivel individual) y las asociaciones observadas a nivel individual no siempre se hallan también a nivel grupal. Por ejemplo, en cada país los mayores ingresos a nivel individual pueden asociarse con una disminución de la mortalidad por enfermedades coronarias, mientras que a nivel de país, el aumento del ingreso per cápita puede asociarse con un aumento (en vez de una disminución) de las tasas de mortalidad por enfermedad coronaria. Siendo así, el uso de datos individuales para inferir asociaciones grupales puede llevar a inferencias incorrectas. Por supuesto, el grado de sesgo con que las asociaciones medidas a nivel grupal estiman las asociaciones a nivel individual (y viceversa) difiere en cada problema

concreto. En diversas publicaciones de salud pública (34, 61-63) se han discutido las muchas circunstancias que pueden dar lugar a discordancia entre las asociaciones medidas a nivel ecológico y a nivel individual, y las explicaciones matemáticas correspondientes.

En última instancia, las falacias inferenciales que se han comentado son problemas metodológicos que pueden superarse cerciorándose de que los datos de los que se parte corresponden al nivel en que se desea hacer inferencias. En algunos casos, se pueden usar métodos estadísticos para reducir en parte el potencial sesgo resultante de inferir asociaciones de un nivel a otro (62). Más sustancial, sin embargo, es el problema de saber si un estudio que enfoca un solo nivel ignora alguna información que es crucial para entender el problema investigado, por ejemplo, si pueden entenderse adecuadamente los mecanismos que operan a un nivel sin la referencia a otros niveles. Como ha explicado Riley (60), aunque el nivel al que se recolectan los datos puede encajar en el modelo conceptual investigado, pueden pasarse por alto datos importantes correspondientes a otros niveles (con lo que el método no se ajusta a la realidad). Ignorar las variables de nivel grupal pertinentes al estudio de asociaciones a nivel individual puede llevar a lo que Riley llamó "falacia psicologista", que consiste en suponer que los resultados individuales pueden explicarse exclusivamente a partir de características individuales. Un ejemplo hipotético sería el de un estudio basado en individuos en el que se constatará que los inmigrantes tienden más a deprimirse que los nativos. Pero supongamos que esto sea sólo verdad en inmigrantes que viven en comunidades en las que constituyen una minoría pequeña. Un investigador que ignora el efecto contextual de la composición comunitaria podría atribuir las mayores tasas de depresión en inmigrantes a los efectos psicológicos producidos por la inmigración *per se*, o incluso a factores genéticos, ignorando la importancia de factores de nivel comunitario y cometiendo así la falacia psicologista (57, 60). (El término "falacia psicologista" no es demasiado apropiado, porque los factores de nivel individual utilizados para explicar los resultados no necesariamente son psicológicos. Otros autores hablan de "falacia individualista" (57), pero como también se ha usado ese término como sinónimo de "falacia atomista" (22, 64), aquí no se utilizará.)

Análogamente, si no se tienen en cuenta los factores del nivel individual en un estudio grupal se puede llegar a la llamada "falacia sociologista" (60). Riley proporciona un ejemplo relacionado con la salud pública. Supongamos que un investigador halla mayores tasas de suicidio en las comunidades con mayor proporción de residentes temporales y concluye que una proporción elevada de población transeúnte lleva a la desorganización social, a la ruptura de las redes sociales y aumenta el riesgo de suicidio en el conjunto de los habitantes de la comunidad. Es posible que los datos individuales muestren que la mayoría de los suicidios ocurre en residentes temporales y que la tasa de suicidio en residentes temporales y en residentes permanentes no varíen de unas comunidades a otras. Si es éste el caso, el investigador estaría cometiendo la falacia sociologista al atribuir la tasa de suicidio elevada a la desorganización social que afectaría a todos los miembros de la comunidad, en lugar de atribuirla a las diferencias intergrupales de proporción de residentes temporales que pueden ser más proclives al suicidio debido a factores de nivel individual. Se puede pensar las falacias psicologista y sociologista como tipos especiales del fenómeno de confusión. En ambos casos se omiten variables pertinentes que pertenecen a otros niveles del modelo. Al incorporar múltiples niveles de

determinación, el análisis multinivel de resultados individuales permite estimar efectos de variables de nivel macro y micro, así como sus interacciones, evitando tanto la falacia psicologista como la sociologista.

Ambos grupos de falacias (ecológica y atomística frente a sociologista y psicologista) están relacionadas. Cuando una variable grupal se relaciona con el resultado independientemente de la misma variable medida a nivel individual, o cuando la variable grupal modifica los efectos de la variable individual en el resultado (es decir, cuando existen efectos contextuales), los coeficientes de regresión ecológicos diferirán de los coeficientes correspondientes a nivel individual (62, 65, 66). Esto ocurre porque los efectos del nivel individual y grupal de la variable están confundidos en el coeficiente de regresión ecológico. Como ilustraron Greenland y Robins (62), la existencia de efectos contextuales es una fuente (aunque hay otras) de falacia ecológica. Los diferentes tipos de falacias se resumen esquemáticamente en la Tabla 2.

### *Otros aspectos relevantes del análisis multinivel o contextual*

Una de las críticas principales al análisis contextual se refiere a lo que se ha llamado "la especificación incorrecta del modelo a nivel individual", que consiste en que el "efecto contextual o de grupo" observado puede ser debido a la omisión de variables del nivel individual que se relacionan a la vez con el resultado y con la característica grupal investigada (21, 67, 68). Hay aquí básicamente un fenómeno de confusión (o de confusión residual), que es frecuente en estudios epidemiológicos, incluso los basados exclusivamente en datos de nivel individual. Por ejemplo, supongamos que el nivel de violencia vecinal (medido por el número medio de crímenes violentos cometido en el vecindario cada año) se comprueba asociado con el aumento del riesgo de hipertensión, una vez ajustado el efecto de la edad y el género. Estos resultados podrían considerarse sugerentes de que la violencia en el vecindario puede relacionarse con el desarrollo de hipertensión, posiblemente a través de sus efectos en los niveles de estrés experimentados por los individuos. Por otro lado, también es posible que variables pertinentes del nivel individual se hayan omitido en el modelo y que los efectos del vecindario observados se deban a un fenómeno de confusión. Si las personas que viven en vecindarios más violentos tienden a ser de bajos ingresos y las personas de bajo ingreso tienen alto riesgo de hipertensión (debido a las diferencias de alimentación relacionadas con nivel de ingreso, obesidad u otros factores), el efecto del vecindario desaparecería al incluir en el modelo el ingreso individual.

Ha habido considerable debate sobre la cuestión del fenómeno de confusión debida a variables de nivel individual en el análisis multinivel (67-71). Cuando la variable de nivel individual omitida está relacionada con el resultado que se analiza y no es una variable intermediaria en la cadena causal que vincula características grupales con la enfermedad, su omisión llevará claramente a una estimación errónea del efecto grupal. Sin embargo, si la variable de nivel individual forma parte de la cadena causal que va desde las variables de nivel grupal al resultado analizado, el efecto grupal podrá estimarse mejor a partir de un modelo no



ajustado respecto de dicha variable. Obviamente, para tener influencia sobre la salud el "contexto" debe "entrar en el cuerpo" de algún modo, operando a la postre a través de procesos a nivel individual; por lo tanto, si se incluyeran suficientes variables de nivel individual en los modelos de regresión, el efecto a nivel grupal desaparecería. No obstante, esto no implica que las variables individuales y no las grupales sean las causas "verdaderas" del resultado. Un problema análogo aparece en estudios basados en datos de nivel individual cuando se incluyen en el modelo de regresión variables intermediarias de la cadena causal, que lleva de la exposición a la enfermedad. En última instancia decidir si una variable dada es una causa independiente o un paso intermedio en una cadena causal que implica variables menos próximas no es algo que pueda resolverse exclusivamente por métodos estadísticos. Hay que estudiar en concreto los mecanismos causales implicados.

Además del fenómeno de confusión debida a variables de nivel individual, el análisis multinivel tiene otra dimensión que también implica posibles fenómenos de confusión, que pueden darse a nivel grupal si hubo variables grupales pertinentes omitidas en el modelo (21). En el ejemplo anterior, el aumento de la violencia en un vecindario podría estar asociado con otra variable grupal, la debilidad de las redes sociales y la menor cohesión social que podrían ser el factor grupal pertinente realmente relacionado con la hipertensión.

Igual que ocurre con el fenómeno de confusión, la multicolinealidad entre variables no es específica del análisis multinivel y a menudo está presente en los estudios en los que sólo se usan datos de nivel individual. En el análisis multinivel también puede haber multicolinealidad entre variables grupales, y entre variables grupales e individuales. En algunos casos la correlación entre variables de nivel grupal (por ejemplo, el nivel medio de ingreso y la tasa de desempleo en el vecindario), pueden ser substanciales. En grupos pequeños y homogéneos, las variables derivadas (construidas por agregación de las características individuales) pueden estar muy correlacionadas con sus análogos a nivel individual (por ejemplo, el ingreso medio del vecindario y el ingreso individual) (21). Si dos variables son intensamente colineales, puede ser sumamente difícil (y quizás muy poco significativo) descomponer sus efectos independientes. Por supuesto, en cada problema concreto puede variar mucho el grado en que se presentan diversos tipos de multicolinealidad.

El análisis contextual también plantea dos problemas de selección, la selección de la unidad contextual apropiada y la de las variables contextuales que deben incluirse en el análisis. Ambas dependen fundamentalmente de cuál es la pregunta específica de la investigación y del modelo teórico del que se parte. Seleccionar la unidad contextual apropiada puede ser problemático porque en general los contextos no están claramente definidos y suelen tener límites imprecisos. Por ejemplo, si el vecindario es la unidad contextual de interés, los investigadores necesitan especificar exactamente en términos operativos qué constituye un vecindario. Los límites de un vecindario y cómo lo perciben sus habitantes no siempre coinciden con las unidades geográficas (por ejemplo áreas censales o distritos municipales) para las que hay datos. Por supuesto, muchos "contextos" pertinentes para la investigación sanitaria (como los lugares de trabajo o las escuelas) pueden no tener nada que ver con un área geográfica de residencia; otros (organizaciones profesionales, redes de familiares), pueden no estar geográficamente definidos. Además, los individuos son parte de diversos contextos superpuestos

(21) cuyos efectos independientes pueden ser sumamente difíciles de separar. Por ejemplo, si un vecindario está fuertemente segregado en cuanto al tipo de trabajo de los residentes, los efectos del contexto vecinal y laboral pueden confundirse y será difícil aislarlos. Seleccionar las variables contextuales apropiadas requiere delimitar el constructo que se desea medir y decidir qué variables se usarán para medirlo. A menudo, incluso cuando está claro el constructo grupal pertinente, las variables necesarias para medirlo pueden no estar disponibles y habrá que usar indicadores indirectos.

Una cuestión metodológica importante que puede surgir cuando el estudio incluye datos de múltiples niveles es la posibilidad de que los resultados individuales intragrupal que se analizan puedan estar correlacionados incluso una vez ajustado el efecto de las variables individuales y grupales. Por ejemplo, en un estudio del efecto de los factores de nivel escolar en el uso de drogas por los adolescentes, en el que se toman muestras de 20 adolescentes por escuela en 100 escuelas, los datos de los alumnos de una escuela pueden estar correlacionados incluso una vez ajustados los efectos de las variables de nivel individual y grupal (porque los alumnos de una misma escuela pueden compartir otras variables no incluidas en el análisis). Esta correlación residual viola el supuesto de independencia de las observaciones que es habitual en los modelos estadísticos usuales. Para resolver el problema planteado por esa correlación, que es análoga a la que surge en el análisis de datos longitudinales de mediciones repetidas en un mismo individuo, se han desarrollado diversas estrategias como los modelos jerárquicos lineales o modelos de efectos aleatorios (23, 28, 53, 72-76) y los modelos marginales o de promedio poblacional (77). Hay muchas publicaciones que tratan los detalles estadísticos y supuestos de estos modelos, así como los casos en que las inferencias obtenidas a partir de ellos difieren de las obtenidas a partir de los modelos de regresión habituales.

## *Conclusión*

Quizás el aspecto más difícil y más interesante del análisis multinivel es que requiere una teoría causal que integre variables de los niveles micro y macro y explique las relaciones e interacciones entre niveles. ¿Cómo se supone que operan las variables de nivel grupal? ¿Cómo interactúan los individuos con sus contextos? Puede ser que estos nuevos modelos exijan ir más allá de la noción simplista de causalidad de la que se parte en muchas investigaciones epidemiológicas (11), al incorporar otros niveles de determinación, como por ejemplo la determinación estructural u holística (78). Tal como la define Bunge, la determinación estructural se refiere al "proceso por el cual la conducta de un individuo (una molécula de un fluido, una persona en un grupo social) es determinada por la estructura global de la colectividad a la que pertenece" (78, p.19). Loomis y Wing (79) aluden a una noción similar cuando se refieren a la necesidad de considerar una causa "no como una propiedad de ciertos agentes sino de los sistemas en los que ocurren los fenómenos de salud y enfermedad" y conciben las poblaciones como grupos organizados con propiedades relacionales, no como simples agregados de individuos. La complejidad de las formulaciones teóricas que relacionan múltiples niveles es una dificultad

importante en el análisis multinivel (22), pero esa complejidad probablemente refleja mejor la realidad que el modelo multicausal más simple que hoy prevalece.

El análisis multinivel no es fácil. Plantea numerosos problemas teóricos y metodológicos que todavía están por resolver. Además, hay que subrayar que la inclusión de variables contextuales no es indispensable para investigar los determinantes sociales de la salud. De hecho, muchas hipótesis que consideran las influencias sociales sobre la salud pueden y deben someterse a contrastación empírica usando datos de nivel individual. Sin embargo, como los procesos causales de enfermedad comprenden diversos niveles y probablemente involucran la interacción de variables de nivel individual y de nivel macro (14, 79), parece apropiado comenzar a desarrollar modelos etiológicos que integren esos niveles para incluir así ambos tipos de variables en los estudios epidemiológicos.

Paradójicamente, la epidemiología, que es el estudio de la enfermedad en poblaciones, ha quedado reducida en gran medida a estudiar los factores de riesgo de enfermedad a nivel individual. El análisis multinivel es una manera de comenzar a reincorporar una dimensión poblacional o social a la investigación epidemiológica: la idea de que factores que operan a nivel de grupos o sociedades afectan la salud de los individuos dentro de esos grupos exige a los epidemiólogos desarrollar modelos causales que integren determinantes de nivel macro y micro. Junto con otras estrategias analíticas y diseños de estudio, el análisis multinivel puede contribuir a fortalecer y revitalizar la investigación de los determinantes sociales y colectivos de la salud.

**Tabla N° 1: Tipos de estudios según los niveles a los que corresponden las variables independientes y la variable dependiente**

VARIABLES INDEPENDIENTES	VARIABLE DEPENDIENTE	TIPO DE ESTUDIO
GRUPAL	Grupal	Ecológico
INDIVIDUAL	Individual	Individual
GRUPAL E INDIVIDUAL	Individual	Multinivel o contextual

**Tabla N° 2: Tipos de falacias**

UNIDAD DE ANÁLISIS	NIVEL DE INFERENCIA	TIPO DE FALACIA
GRUPAL	Individual	Ecológica
INDIVIDUAL	Grupal	Atomista*
INDIVIDUAL (excluidas las variables de nivel grupal pertinentes)	Individual	Psicologista*
GRUPAL (excluidas las variables de nivel individual pertinentes)	Grupal	Sociologista

\* También llamada falacia individualista por algunos autores.

## Reconocimientos

Agradezco a Carlos Muntaner, Javier Nieto y José Tapia sus muchas contribuciones a las ideas elaboradas en este artículo.

*Traducción: José Tapia*

## Bibliografía

1. Cassel J. Social science theory as a source of hypotheses in epidemiological research. *Am. J. Public Health* 1964;54:1482-88.
2. Susser M, Susser E. Choosing a future for epidemiology I: Eras and paradigms. *Am. J. Public Health* 1996;86:668-673.
3. Pearce N. Traditional epidemiology, modern epidemiology, and public health. *Am. J. Public Health* 1996;86:678-683.
4. Catalano R. Paradigm succession in the study of public health. En: *Health, behavior, and the community*. New York: Pergamon Press; 1979:87-137.
5. Brockington CE The history of public health. En: Hobson W. ed. *Theory and practice of public health*. 5th edition. London: Oxford University Press; 1979:1-8.
6. Taylor R, Rieger A. Medicine as social science: Rudolph Virchow on the typhus epidemic in upper Silesia. *Int. J. Health Serv.* 1985;15:547-559.
7. Fee E. The origins and development of public health in the United States. En: Holland W., Deteb R., Knox G. eds. *The Oxford Textbook of Public Health* 2ed. vol. 1. London: Oxford University Press; 1991:3-22.
8. Kennedy DA. Community health and the urban environment. En: Hinkle LE., Loring WC. eds. *The effect of the man made environment on health and behavior*. Atlanta: Centers for Disease Control, Public Health Service 1977. DHEW Public N°. 77-8318. 1977:7-44.
9. Loring WC. Introduction. En: Hinkle LE., Loring WC. eds. *The effect of the man made environment on health and behavior*. Atlanta: Centers for Disease Control. US Dept of Health, Education, and Welfare. DHEW Public N°. (CDC) 77-8318. 1977:vii-xxxvi.
10. Dodge DL, Martin WT. The root of the problem: changing causes of death. En: *Social stress and chronic illness*. Notre Dame: University of Notre Dame Press; 1970:3-25.
11. Krieger N. Epidemiology and the web of causation: has anyone seen the spider? *Soc. Sci. Med.* 1994;39:887-903.
12. Duncan C, Jones K, Moon G. Health-related behaviour in context: a multilevel modelling approach. *Soc. Sci. Med.* 1996;42:817-830.
13. Lukes S. Methodological individualism reconsidered. En: Emmet D., Macintyre A. eds. *Sociological theory and philosophical analysis*. New York: Macmillan; 1970:76-88.
14. Von Korff M, Koepsell T, Curry S, Diehr P. Multi-level research in epidemiologic research on health behaviors and outcomes. *Am. J. Epidemiol.* 1992;135:1077-1082.
15. Susser M. The logic in ecological: I. The logic of analysis. *Am. J. Public Health* 1994;84:825-829.
16. Krieger N, Rowley DL, Hermán A, et al. Racism, sexism, and social class: implica-

- tions for studies of health, disease, and well-being. *Am. J. Prev. Med.* 1993;9(6 suppl):82-122.
17. Jones K, Duncan C. Individuáis and their ecologies: analysing the geography of chronic illness within a multilevel modelling framework. *Health & Place* 1995;1:27-40.
  18. Blalock HM, Wilken PH. *Integroup processes. A micro-macro perspective.* New York: The Free Press; 1979:1-23 288-323.
  19. Dogan M, Rokkam S. Introduction. En: Dogan M., Rokkam S. eds. *Social Ecology* Boston: The MIT Press; 1969:1-15.
  20. Iversen G. *Contextual analysis.* Newbury Park CA: Sage Publications; 1991.
  21. Blalock HM. Contextual-effects models: theoretical and methodological issues. *AnnRev Sociol* 1984;10:353-372.
  22. Scheuch EK. Social context and individual behavior. En: Dogan M., Rokkam S. eds. *Social Ecology* Boston: The MIT Press; 1969:133-155.
  23. Hermalin AL The multilevel approach: theory and concepts. En: *The methodology for measuring the impact of family planning programmes on fertility.* Population Studies No. 66. Addendum Manual IX. New York: United Nations; 1986: 15-31.
  24. Hox J, Kreft I. Multilevel analysis methods. *Sociological Methods Res.* 1994;22:283-299.
  25. Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. *Annu. Rev. Public Health* 1995;16:61-81.
  26. Lazarsfeld PF, Menzel H. On the relation between individual and collective properties. En: Etzioni A, ed. *A sociological reader on complex organizations.* New York: Holt, Rinehart, and Winston.;1971:499-516.
  27. Przeworski A. Contextual models of political behavior. *Political methodology* 1974;1:27-61.
  28. Garner C, Raudenbush S. Neighborhood effects on educational attainment: a multilevel analysis. *Sociology of Education* 1991;64:251-262.
  29. Robson BT. *Urban analysis: a study of city structure with special reference to Sutherland.* Cambridge: Cambridge University Press; 1969.
  30. Fernandez RM, Kulik JC (1981). A multilevel model of life satisfaction: effects of individual characteristics and neighborhood composition. *Am. Sociol. Rev.* 1981;46:840-850.
  31. Morgenstern H. Socioeconomic factors: concepts, measurement, and health effects. En: *Measuring psychosocial variables in epidemiologic studies of cardiovascular disease: proceedings of a work-shop.* Washington DC: National Institutes of Health. NIH publication no. 85-2270. 1985:3-35.
  32. Wing S, Casper M, Riggan W, et al. Socioenvironmental characteristics associated with the onset of decline of ischemic heart disease mortality in the United States. *Am. J. Public Health* 1988;78:923-926.
  33. Wing S, Barnett E, Casper M, Tyroler HA. Geographic and socioeconomic variation in the onset of decline of coronary heart disease mortality in white women. *Am. J. Public Health* 1992;82:204-209.
  34. Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and its consequences. *Am. J. Public Health* 1994;84:819-824.
  35. Susser M, Susser E. Choosing a future for epidemiology I: From black box to Chi-nese boxees and ecoepidemiology. *Am. J. Public Health* 1996;86:674-677.

36. Goldberger J, Wheeler GA, Sydenstrycker E. A study of the relation of family in-come and other economic factors to pellagra incidence in seven cotton mili villa-ges of South Carolina in 1916. *Pub. Health Rep.* 1920;35:2673-2714.
37. Macintyre S, Maciver S, Sooman A. Área, class and health: should we be focusing on places or people? *Jnl. Soc. Pol.* 1993;22:213-234.
38. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation* 1993;88:1973-1998.
39. Carstairs V, Morris R. Deprivation and mortality. An alternative to social class? *Comm. Med.* 1989;11:210-219.
40. Carstairs V, Morris R. Deprivation: explaining differences in mortality between Scotland, England and Wales. *Br. Med. J.* 1989;299:886-889.
41. Morgan M, Chinn S. ACORN group, social class and child health. *J. Epidemiol. Comm. Health* 1983;37:196-203.
42. Fox AJ, Jones DR, Goldblatt PO. Approaches to studying the effect of socioeconomic circumstances on geographic differences in mortality in England and Wales. *Br. Med. Bull* 1984;4:309-314.
43. Haan M, Kaplan G, y Camacho T Poverty and health: prospective evidence from the Alameda County Study. *Am. J. Epidemiol.* 1987;125:989-998.
44. Hochstim JR. Poverty área under the microscope. *Am. J. Public Health* 1968; 58:1823-1827.
45. Anderson RT, Sorlie P, Backlund E, Johnson N, Kaplan GA. Mortality effects of community socioeconomic status. *Epidemiology* 1997;8:42-47.
46. Humphreys K, Carr-Hill R. Área variations in health outcomes: artefact or ecóloga. *Int. J. Epidemiol.* 1991;20:251-258.
47. Krieger N. Women and social class: a methodological study comparing individual, household, and census measures as predictors of black/white differences in repro-ductive history. *J. Epidemiol. Comm. Health* 1991;45:35-42.
48. Krieger N. Overcoming the absence of socioeconomic data in medical records: va-lidation and application of a census-based methodology. *Am. J. Public Health* 1992; 92:703-710.
49. Diehr P, Koepsell T, Cheadle A, Psaty BM, Wagner E, Curry S. Do communities differ in health behaviors? *J. Clin. Epidemiol.* 1993;46:1141-1149.
50. Duncan C, Jones K, Moon G. Do places matter? A multi-level analysis of regional variation in health-related behavior in Britain. *Soc. Sci. Med.* 1993;37:725-733.
51. Duncan C, Jones K, Moon G. Psychiatric morbidity: a multilevel approach to regional variations in the UK. *J. Epidemiol. Comm. Health* 1995;49:290-95.
52. O'Campo P, Gielen AC, Faden RR, Xue X, Kass N, Wang M. Violence by male partners against women during the childbearing year: a contextual analysis. *Am. J. Public Health* 1995;85:1092-1097.
53. Entwisle B, Masón WM. Multilevel effects of socioeconomic development and fa-mily planning programs on children ever born. *Am. J. Sociol.* 1985;91:616-649.
54. Dargent-Molina P, James SA, Strogatz DS, Savitz DA. Association between mater-nal education and infant diarrhea in differenr households and community envi-ronments of Cebú, Philippines. *Soc. Sci. Med.* 1994;38:343-350.
55. Koopman JS, Longini IM. The ecological effects of individual exposures and non-linear disease dynamics in populations. *Am. J. Public Health* 1994;84:836-842.
56. Jones K, Moon G, Clegg A. Ecological and individual effects in childhood immuni-

- sation uptake: a multilevel approach. *Soc. Sci. Med.* 1991;33:501-508.
57. Valkonen T. Individual and structural effects in ecological research. En: Dogan M, Rokkam S, eds. *Social Ecology*. Boston: The MIT Press; 1969:53-68.
  58. Koopman JS, Longini IM, Jacques JA, Simón CP, Ostrow DG, Martin WR, Wood-cock DM. Assessing risk factors for transmission of infection. *Am. J. Epidemiol.* 1991;1199-1209.
  59. Halloran ME, Struchiner CJ. Study design for dependent happenings. *Epidemiology* 1991;2:331-338.
  60. Riley MW. Special problems of sociological analysis. En: *Sociological Research I: A case approach*. New York: Harcourt, Brace, and World; 1963:700-725.
  61. Piantadosi S, Byar DP, Green SN. The ecological fallacy. *Am. J. Epidemiol.* 1988;127
  62. Greenland S, Robins S. Ecologic studies- biases, misconceptions, and counter-examples. *Am. J. Epidemiol.* 1994;139:747-760.
  63. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int. J. Epidemiol.* 1989;18:269-274.
  64. Alker HR. A typology of ecological fallacies. En: Dogan M., Rokkam S. eds. *Social Ecology*. Boston: The MIT Press; 1969:69-86.
  65. Hammond JL. Two sources of error in ecological correlations. *Am. Sociol. Rev.* 1973;38
  66. Firebaugh G. A rule for inferring individual-level relationships from aggregate data. *Am. Sociol. Rev.* 1978;43:557-572.
  67. Hauser RM. Contextual analysis revisited. *SocioL Meth. Res.* 1974;2:365-375.
  68. Hauser RM. Context and consex: a cautionary tale. *Am. J. SocioL* 1970;75:645-664.
  69. Hauser RM. Reply to Alien Barton. *Am. J. SocioL* 1970;76:517-520.
  70. Barton A. Comments on Hauser's "Context and consex". *Am. J. SocioL* 1970; 76: 514-517.
  71. Farkas G. Specification, residuals, and contextual effects. *Sociológica! Methods Res.* 1974;2:333-363.
  72. Bryk A, Raudenbush S. *Hierarchical linear models: applications and data analysis methods*. Mewbury Park CA: Sage Publications; 1992.
  73. Wong GY, Masón WM. The hierarchical logistic regression model for multilevel analysis. *J. Am. Stat Assoc* 1985;80:513-524.
  74. Wong GY, Masón WM. Contextually specific effects and other generalizations of the hierarchical linear model for comparative analysis. *JASA* 1991;86:487-503.
  75. Entwisle B, Masón WM, Hermal HI. The multilevel dependence of contraceptive use on socioeconomic development and family planning program strength. *Demo-graphy* 1986;23:199-216.
  76. Masón WM, Wong GW, Entwisle B. Contextual analysis through the multilevel linear model. En: Leinhardt S. ed. *Sociological Methodology 1983-1984*. San Francisco: Josey Bass; 1983:72-103.
  77. Zeger S, Liang KY, Albert P. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988;44:1049-1060.
  78. Bunge M. Causation and determination, causalism and determinism. En: *Causality in Modern Science*. New York: Dover Publications; 1979:3-30.
  79. Loomis D, Wing S. Is molecular epidemiology a germ theory for the end of the twentieth century? *Int. J. Epidemiol.* 1990;19:1-3.