

**NIVELL SOCIOECONÒMIC I PREVALÈNÇA D'ASMA EN  
ADULTS JOVES: UN ANÀLISI MULTINIVELL EN  
L'EUROPEAN COMMUNITY RESPIRATORY HEALTH  
SURVEY**

**Autor: Xavier Basagaña Flores  
Director: Josep Ginebra**

**Abril de 2003**

# Agraïments

Primer de tot he d'agrair a en Josep Ginebra els seus consells i comentaris sobre el projecte que sens dubte l'han fet millorar molt, i al Dr. Jordi Sunyer de l'IMIM que ha estat la persona amb qui hem dissenyat i realitzat aquest anàlisi, i també a l'altra gent de l'IMIM que hi ha participat aportant comentaris.

També vull agrair des d'aquí tota l'ajuda desinteressada que m'han donat en Josep Ginebra, la Lupe Gómez i l'Erik Cobo, qui, a més de ser grans professors m'han ajudat molt, moltíssim, en coses no directament relacionades amb la Universitat. Que hi hagi gent com vosaltres disposada a ajudar fa que les coses siguin molt més fàcils. Espero jo poder seguir el vostre exemple.

També als meus pares, perquè sense ells res del que m'està passant hagués estat possible. Les coses que considero més importants les he après de vosaltres.

També al José Miguel per estar sempre disposat a ajudar-me i contagiar-me les seves ganes d'aprendre.

I finalment a la Mercè, per tot el que ja m'ha donat, però sobretot per tot el que encara ha de venir.

# Contingut

<b>CAPÍTOL 1: INTRODUCCIÓ.....</b>	<b>1</b>
1.1. PRESENTACIÓ DEL PROBLEMA .....	1
1.2. RESUM DE LES ANÀLISIS ESTADÍSTIQUES EFECTUADES .....	2
1.3. RESUM DE LES CONCLUSIONS .....	4
<b>CAPÍTOL 2: DESCRIPCIÓ DE L'ESTUDI ECRHS.....</b>	<b>5</b>
2.1 INTRODUCCIÓ.....	5
2.2 DEFINICIONS.....	7
2.2.1. <i>Asma:</i> .....	7
2.2.2. <i>Al·lèrgia: Hipersensibilitat cutània i IgE sèrica</i> .....	10
2.2.2.1 Elecció de les proves .....	10
2.2.2.2 Elecció dels al·lèrgens.....	11
2.3 METODOLOGIA DE L'ESTUDI .....	12
<b>CAPÍTOL 3. ASMA I NIVELL SOCIOECONÒMIC.....</b>	<b>14</b>
3.1. INTRODUCCIÓ .....	14
3.2. ANTECEDENTS.....	14
3.3. OBJECTIU.....	17
3.4 VARIABLES A NIVELL INDIVIDUAL.....	17
3.4.1 <i>Variable resposta</i> .....	17
3.4.2 <i>Variables socioeconòmiques</i> .....	18
3.4.3 <i>Variables confusores</i> .....	19
3.5. VARIABLES A NIVELL DE CENTRE O ECOLÒGIQUES.....	22
3.5.1. <i>Variables de classe social</i> .....	22
3.5.2. <i>Variables confusores</i> .....	22
<b>CAPÍTOL 4. DESCRIPCIÓ DELS MODELS I LA SEVA INTERPRETACIÓ .....</b>	<b>24</b>
4. 1. INTRODUCCIÓ.....	24
4.2. QUAN UTILITZAR EFECTES ALEATORIS .....	26
4.3. MODELS LINEALS MIXTES .....	28

4.3.1. Model amb la constant aleatòria.....	28
4.3.2. Models amb pendents aleatoris .....	29
4.3.3 Introducció de variables de nivell superior.....	30
4.3. MODELS LOGÍSTICS MIXTES .....	32
4.4. INTERPRETACIÓ DELS EFECTES ALEATORIS. ....	33
4.5. MÈTODES D'ESTIMACIÓ I SOFTWARE .....	35
<b>CAPÍTOL 5. ESTRATÈGIES DE MODELITZACIÓ, VALIDACIÓ I INTERPRETACIÓ .....</b>	<b>38</b>
5.1. ESTRATÈGIA D'ANÀLISI .....	38
5.1.1. Anàlisi exploratòria de dades.....	38
5.1.2. Modelització .....	41
5.1.2.1. Estratègia escollida.....	41
5.1.2.2. Alternatives .....	46
5.2. ESTRATÈGIA DE VALIDACIÓ .....	47
5.2.1. Residus a nivell individual.....	48
5.2.2. Residus a nivell de centre.....	50
5.3. ESTRATÈGIA D'INTERPRETACIÓ.....	52
5.3.1. Interpretació dels resultats.....	52
5.3.2. Interpretació dels odds ratio .....	53
<b>CAPÍTOL 6. ANÀLISI EXPLORATÒRIA DE LES DADES.....</b>	<b>56</b>
6.1 DESCRIPCIÓ DE LES VARIABLES INDIVIDUALS PER CENTRE .....	56
6.1.1. Participació i variable asma .....	56
6.1.2 Variables socioeconòmiques .....	60
6.1.3 Variables confusores .....	63
6.2. DESCRIPCIÓ DE LES VARIABLES A NIVELL DE CENTRE.....	67
6.3. ASSOCIACIONS CRUES DE LES VARIABLES INDIVIDUALS AMB ASMA.....	68
6.3.1. Variables socioeconòmiques .....	68
6.3.2. Variables confusores .....	69
6.4. ASSOCIACIONS CRUES DE LES POSSIBLES CONFUSORES I EL NIVELL SOCIOECONÒMIC ...	71
6.5. ASSOCIACIONS DE LES VARIABLES A NIVELL DE CENTRE I ASMA .....	75
<b>CAPÍTOL 7. SELECCIÓ DE MODELS .....</b>	<b>78</b>
7.1. DESCRIPCIÓ .....	78
7.2. VALIDACIÓ .....	81
7.2.1 Residus a nivell d'individu.....	81

7. 2. 2 Residus a nivell de centre .....	82
7.3. INTERPRETACIÓ .....	87
7.3.1. Variables socioeconòmiques .....	87
7.3.1.1. Nivell d'individu .....	87
7.3.1.2. Nivell de centre .....	88
7.3.2. Variables confusores .....	89
<b>CAPÍTOL 8. CONCLUSIONS.....</b>	<b>90</b>
8.1. DISCUSSIÓ DE LES TROBALLES.....	90
8.1.1. Resum dels principals resultats .....	90
8.1.2. Aportacions de l'anàlisi multinivell .....	91
8.1.3. Consistència amb altres estudis .....	91
8.1.4. Interpretació de les mesures socioeconòmiques.....	92
8.1.5. Possibles explicacions a l'associació.....	93
8.1.6. Punts forts i punts febles de l'estudi.....	95
8.2. CONCLUSIÓ FINAL .....	97
8.3. TREBALL FUTUR.....	97
<b>BIBLIOGRAFIA REFERENCIADA.....</b>	<b>99</b>
<b>ANNEX 1: ARTICLES DE L'ECRHS LIDERATS PEL CENTRE DE BARCELONA.....</b>	<b>106</b>

# Capítol 1: Introducció

## 1.1. Presentació del problema

El present projecte està realitzat a l'Institut Municipal d'Investigació Mèdica (IMIM), centre en el qual treballa des de fa més de quatre anys, en concret a la secció d'Epidemiologia, a la Unitat de Recerca Respiratòria i Ambiental sota la direcció del Dr. Jordi Sunyer.

Les dades que es fan servir en aquest projecte són d'un estudi multicèntric europeu, l'*European Community Respiratory Health Survey* (ECRHS), del qual Barcelona n'és un dels centres participants i a més a més és el coordinador dels centres espanyols. A partir d'aquest estudi s'han publicat un gran nombre d'articles, en molts dels quals els membres del meu grup (com el Dr. Jordi Sunyer, Dr. Josep Maria Antó, Dr. Manolis Kogevinas, Dr. Jan-Paul Zock, o Dr. Joan B. Soriano) en són coautors, a més de ser membres la majoria d'ells de l'*Steering Comitee* de l'estudi.

Quan jo vaig entrar a l'IMIM l'estudi ja estava en marxa i ja se n'havien publicat bastants articles. Per tant, les dades ja les vaig trobar entrades i netejades. No obstant, he participat en unes quantes anàlisis amb aquestes dades. Els articles que llisto a continuació són articles de l'estudi ECRHS en els quals he participat i a més en sóc coautor.

Basagana X, Sunyer J, Zock JP, Kogevinas M, Urrutia I, Maldonado JA *et al.* Incidence of asthma and its determinants among adults in Spain. *American Journal of Respiratory & Critical Care Medicine* 2001;164:1133-7.

Jaén A, Sunyer J, Basagaña X, Chinn S, Zock JP, Antó JM, Burney P. Specific sensitization to common allergens and pulmonary function in the European Community Respiratory Health Survey. *Clin Exp Allergy* 2002; 32:1713-9

Sunyer J, Basagana X, Burney P, Anto JM. International assessment of the internal consistency of respiratory symptoms. European Community Respiratory Health Study (ECRHS). *American Journal of Respiratory & Critical Care Medicine* 2000;162:930-5.

Actualment s'està fent un seguiment de la gent que va participar a l'estudi, a l'inici dels anys 90, per fer-ne una segona part, l'ECRHS-II, que ara mateix està en la fase de depuració de les dades, i en aquesta segona part sí que m'ocupo de la gestió de dades dels centres espanyols.

En aquest projecte final de carrera es presenta una part d'un estudi que encara no ha estat publicat que investiga la relació entre asma i nivell socioeconòmic, és a dir, si, per exemple, són les classes socials més baixes o més pobres les que pateixen més asma. Aquest aspecte encara no havia estat tocat per l'estudi ECRHS, i altres estudis que han investigat el tema no arriben a conclusions clares. A l'IMIM estem treballant actualment en la redacció d'un article sobre això, que inclourà part de les anàlisis que es presenten en aquest projecte final de carrera però que abarcarà també una diferenciació dels diversos tipus d'asma, com per exemple l'asma al·lèrgic o el no al·lèrgic. Aquestes anàlisis addicionals no estan inclosos en aquest projecte final de carrera perquè augmentarien molt la quantitat de treball i allargarien la finalització del projecte, i a canvi no aportarien cap novetat en termes d'anàlisi ja que la metodologia que es fa servir és la mateixa però fent servir altres variables resposta.

## **1.2. Resum de les anàlisis estadístiques efectuades**

L'anàlisi en qüestió utilitza dades de 10,971 individus repartits en 32 centres que pertanyen a 15 països diferents. En l'anàlisi es fan servir variables socioeconòmiques tant a nivell individual com a nivell de centre. A nivell individual s'utilitza la classe social a la que pertany l'individu i l'educació màxima que ha assolit, i a nivell de centre s'utilitzen variables com el Gross National Product (GNP) del país o el percentatge

d'individus que tenen estudis superiors en cada centre. La relació entre asma i variables socioeconòmiques ha estat ajustada per tots els factors que ja s'han trobat en estudis previs (de l'ECRHS o d'altres) que poden tenir alguna relació amb l'asma i que per tant poden confondre la relació.

L'anàlisi pretén utilitzar a l'hora les variables que són de nivell d'individu i les que són de nivell de centre, i en conseqüència necessita utilitzar models jeràrquics, coneguts també com a models multinivell o models amb efectes aleatoris. Com que la variable resposta és dicotòmica (presència o no d'asma en l'últim any), cal utilitzar models lineals generalitzats amb efectes aleatoris.

El model lineal amb efectes aleatoris és un model conegut, que està molt estudiat i que fa temps que s'aplica. No obstant, els avenços en models lineals generalitzats amb efectes aleatoris, sobretot pel que fa a software que els estimi, s'han fet molt recentment. Les primeres implementacions apareixen publicades els anys 1984 i 1985. Des de llavors han anat sortint articles sobre el tema i en alguns llibres de models jeràrquics com el Snijders i Bosker (1999) s'hi dedica un capítol. No obstant, alguns aspectes d'aquests models són encara una àrea oberta de recerca. Pel que fa al software, en S-Plus encara no hi ha una funció específica pels models lineals generalitzats amb efectes aleatoris, tot i que disposa d'una versió beta, en SAS existeix un procediment que fa servir un mètode aproximat que pot no ser bo per fer inferència en alguns casos, i s'han desenvolupat paquets específics per models multinivell, com MLwiN, HLM i MIXREG/MIXOR. Els models lineals generalitzats amb efectes aleatoris no s'estudien ni a la Diplomatura ni a la Llicenciatura d'Estadística, de manera que, abans de poder fer l'anàlisi, he hagut de dur a terme una recerca d'informació i una preparació per entendre i aplicar adequadament aquests models. Les referències bàsiques que he utilitzat són Snijders i Bosker (1999), Goldstein (1995), Hox (1995), i Diez-Roux (1998, 2000).

De la mateixa manera, tractant-se d'un tema mèdic he hagut d'aprendre sobre les malalties respiratòries, sobretot des de la vessant epidemiològica, i he hagut de realitzar



una cerca bibliogràfica exhaustiva del tema del projecte per conèixer el que s'ha publicat fins ara i avaluar l'estat de la qüestió.

Els resultats que es presenten en aquest projecte són una part dels que s'inclouran en un article que s'està escrivint actualment i que s'enviarà a publicar pròximament a alguna revista epidemiològica.

### **1.3. Resum de les conclusions**

Els resultats del projecte assenyalen que tots els indicadors socioeconòmics a nivell individual mostren que tenen més asma les categories més desfavorides. Així, l'increment de risc que trobem en la classe social més baixa respecte a la més alta és d'un 50%. En l'educació, els que han assolit un menor nivell d'estudis també són els que tenen més asma, encara que la variable educació té un efecte més gran quan es tracta a nivell de centre que no quan es tracta a nivell individual. També observem que aquells que a la seva feina estan exposats a pols mineral o biològica, gas o fums, tenen un increment del risc d'asma d'un 25%. A nivell de centre, també observem una associació que indica que els que viuen en països socioeconòmicament més malestants, tenen un risc afegit d'asma, independentment de les seves característiques socioeconòmiques individuals. En concret, el model final obtingut indica que pertànyer a un centre amb un alt percentatge d'alta educació és protector de cara a tenir asma (per cada 1% més de persones amb educació alta que té el centre, el risc es redueix un 12%).

El fet d'usar models multinivell ha permès avaluar l'efecte de les variables socioeconòmiques a nivell d'individu i a nivell de centre simultàniament, ajustades les unes per les altres. A més a més, ha permès comprovar la homogeneïtat de les relacions per centre. En aquest darrer aspecte hem constatat que els efectes que hem trobat en totes les variables poden considerar-se els mateixos en cada centre. Incloure la variable "percentatge d'alta educació del centre" ha reduït la variància entre centres no explicada un 31%.

# Capítol 2: Descripció de l'estudi

## ECRHS

### 2.1 Introducció

Durant la primera meitat dels 90 es va recollir informació sobre la prevalença<sup>1</sup> d'asma, sobre factors de risc (coneguts i suposats) d'asma i atòpia (caràcter constitutiu de certs individus que presenten reaccions al·lèrgiques amb una freqüència anormalment alta), i informació sobre el tractament de l'asma en adults joves en l'estudi *European Community Respiratory Health Survey* (ECRHS)

La raó per la qual es va portar a terme l'estudi va ser el ràpid increment en la prevalença d'asma que s'havia detectat en diferents països. Aquest increment es va produir en un període molt curt de temps per ser explicat per factors genètics i havia d'estar relacionat amb canvis ambientals. Estudis que comparaven la prevalença en àrees rurals i urbanes en països desenvolupats i trobaven grans diferències anaven a favor de la hipòtesi de les causes ambientals.

Els patrons de tractament de l'asma i la mortalitat per asma varien considerablement a Europa, i això va formular la qüestió de si la variació en la mortalitat era deguda a diferències en la prevalença d'asma o a diferències en la mortalitat per asma. Era necessària una recerca dirigida a identificar els factors ambientals que expliquessin aquesta variabilitat geogràfica en l'asma per identificar estratègies potencials que puguin contrarestar l'augment global en la prevalença d'asma.

---

<sup>1</sup> En Epidemiologia, el percentatge de persones afectades amb una determinada malaltia en un moment donat es coneix com a prevalença

La informació disponible sobre la variació de la prevalença d'asma i al·lèrgia al principi dels 90 s'havia aconseguit gràcies a diversos estudis que usaven diferents protocols, definicions i mètodes. L'estudi ECRHS va ser, doncs, el primer estudi que valorava la prevalença d'asma i al·lèrgia en un gran nombre de països fent servir un protocol idèntic i estandarditzat. Els tres principals objectius de l'estudi eren:

1. Estimar la variació en la prevalença d'asma, símptomes respiratoris semblants a l'asma, sensibilització atòpica i hiperreactivitat bronquial a Europa.
2. Estimar la variació en l'exposició a factors de risc (coneguts o suposats) per l'asma i mesurar la seva associació amb l'asma, i valorar fins a quin punt poden explicar la variació en la prevalença d'asma a Europa.
3. Estimar la variació en el tractament de l'asma a la Comunitat Europea.

Prèviament a la realització de l'estudi, es va acordar un pla de dues etapes per l'anàlisi de les dades, per reduir la probabilitat de reportar associacions degudes a l'atzar que provinguin del fet de testar múltiples hipòtesis. Primer, les dades d'un o més centres d'un país s'analitzarien per examinar potencials factors de risc per la malaltia en qüestió en aquell país. L'estudi d'aquell factor concret només es duu a terme en el país que ho ha sol·licitat o que li ha estat assignat. Un cop aquests resultats han estat publicats, l'anàlisi es repeteix amb la base de dades completa de tot l'estudi.

En l'estudi hi ha participat un gran nombre d'investigadors i treballadors de camp de 22 països. La base de dades de l'ECRHS inclou informació d'aproximadament 140.000 individus, i s'han publicat més de 100 articles sobre diferents aspectes epidemiològics de l'asma amb dades de l'ECRHS. A l'annex 1 es llisten els articles liderats per gent del centre de Barcelona. A part, hi ha hagut nombroses col·laboracions dels membres de Barcelona en altres articles liderats per gent d'altres centres.

L'any 2001 es va publicar un article que resumia els principals resultats que ha aportat l'estudi ECRHS fins a aquell moment (Janson 2001).

## **2.2 Definicions**

### **2.2.1. Asma:**

L'asma és un trastorn respiratori caracteritzat per estrenyiment bronquial contràctil, inflamatori i edematós que origina episodis repetits de dispnea paroxística, sibilàncies espiatòries, tos i secrecions bronquials mucoses viscoses (Perpiñá 1997). Els episodis, reversibles, poden desencadenar-se com a conseqüència de la inhalació d'al·lèrgens o contaminants, infeccions, exercici enèrgic o tensió emocional. Es poden diferenciar dos grans tipus d'asma, com per exemple l'asma al·lèrgica, que és aquella que ve causada per l'exposició de la mucosa bronquial a un antigen inhalat, i l'asma intrínseca, que és aquella que no està causada per l'al·lèrgia. Altres metges i investigadors distingeixen també altres tipus d'asma, com l'asma ocupacional, que és aquella que està relacionada amb un ambient laboral especial, i la infantil, que és la que es manifesta a la infància.

La definició de l'asma segueix sent un problema ja que no se'n coneix l'agent causal primari, l'anatomia patològica és difícil d'obtenir i les manifestacions clíniques són variables i a vegades atípiques. De fet, cap de les iniciatives que s'han portat a terme en els últims 50 anys ha aconseguit una definició acceptada de forma unànime per clínics, epidemiòlegs i fisiopatòlegs. Una de les definicions que sembla plantejar menys problemes és la que diu que « l'asma és un trastorn crònic inflamatori de les vies aèries en la que estan implicats molts elements cel·lulars, particularment mastocits, eosinòfils i limfocits T. En el individu susceptible, aquesta inflamació ocasiona episodis de sibilàncies, dispnea, tibatort toràcica i tos, sobretot a la nit i/o primeres hores de la matinada. Aquests símptomes s'associen, generalment, amb obstrucció generalitzada del flux aeri que, amb freqüència, reverteix espontàniament o amb el tractament. La

inflamació també produeix un augment de la resposta de les vies aèries a estímuls diversos».

Analitzant diversos textos es pot concloure que els elements capitals que engloben el concepte d'asma són, en línies generals:

- a) la inflamació del tracte respiratori;
- b) la hiperresposta bronquial;
- c) l'existència d'àmplies variacions durant curts períodes de temps en la resistència al flux de la via aèria interpulmonar;
- d) la presència d'un acompanyament clínic, generalment de caràcter episòdic i intercalat amb períodes lliures de símptomes, dominat per la dispnea, tos, sibilàncies audibles i tibantor al pit, que cessen ja sigui espontàniament o com a conseqüència de la medicació.

Tot i això, la realitat diària demostra que, tot i utilitzar aquests darrers elements, l'asma no sempre és fàcil de distingir d'altres malalties o disfuncions, ja que les seves manifestacions principals presenten nexes d'unió amb altres malalties amb les que, a més, comparteix alguna de les seves peculiaritats. Per exemple, pel que fa a la *inflamació*, hi ha països subdesenvolupats on tenen baixa prevalença d'asma i en canvi alts nivells d'infeccions de la via aèria. La *hiperresposta bronquial* es dona en altres malalties (bronquitis crònica, fibrosi quística, rinitis al·lèrgica, etc.), i també en individus sans després de determinades infeccions respiratòries o exposicions a contaminació. A més a més, pel que fa a la simptomatologia, hi ha una gran dificultat en decidir com etiquetar a un nen "amb sorolls al pit" o a un adult amb una història de tos i expectoració cròniques a la que se li sumen sibilàncies més o menys persistents.

Aquestes consideracions expressen de forma resumida els greus inconvenients que planteja la definició d'asma i la dificultat per diagnosticar-la i distingir-la, sobretot, de la bronquitis crònica.

En els estudis epidemiològics, la mesura de l'asma es fa seguint criteris operatius, no sempre compartits pels clínics que fan servir criteris difícils d'estandarditzar al comparar cada individu. Així, el que és important en epidemiologia és arribar a definicions que siguin repetibles per tots els subjectes i que tinguin la validesa suficient per identificar amb un mínim d'especificitat (probabilitat que els pacients sans siguin correctament identificats com a sans per la definició) els vertaders asmàtics. Això no només és complex per la pròpia dificultat de mesurar una malaltia en un grup de subjectes i no individualment, sinó que també ho és per les pròpies limitacions del diagnòstic d'asma. En nens, la millor definició de l'asma en estudis epidemiològics és la presentació conjunta de sibilàncies declarades en un qüestionari i la positivitat al test de reactivitat a l'exercici. Tot i això, no sempre és possible mesurar la reactivitat bronquial, per exemple per raons d'edat dels nens o per la no participació als tests de broncoprovocació en els adults sans, i el diagnòstic s'ha de limitar a la declaració de símptomes reportats de manera estàndard en qüestionaris. L'ús de les respostes en qüestionaris respiratoris està acceptat com a mètode vàlid i repetible per mesurar la malaltia respiratòria en estudis poblacionals, si bé a canvi de perdre un cert grau d'especificitat (Samet 1987). Els qüestionaris de l'ECRHS ha estat validats en les diferents llengües (Burney 1994).

La definició que s'ha fet servir és una definició que s'ha usat a tots els articles de l'ECRHS, que és el resultat d'una combinació de preguntes de qüestionari que s'ha mostrat adequada en diversos estudis i ha estat àmpliament usada com a mesura d'asma en l'actualitat. Es considera que una persona pateix asma en l'actualitat si respon afirmativament a alguna de les tres preguntes següents:

- S'ha despertat alguna vegada amb un atac de falta d'aire en els últims 12 mesos?
- Ha tingut algun atac d'asma en els últims 12 mesos?
- Ha pres alguna medicació per l'asma en els últims 12 mesos?

## 2.2.2. Al·lèrgia: Hipersensibilitat cutània i IgE sèrica

L'atòpia és una predisposició genètica per desenvolupar una resposta mitjançant anticossos de la classe IgE a antigens ambientals als que no són sensibles els individus no atòpics. Els atòpics tenen una tendència hereditària a sofrir reaccions al·lèrgiques immediates com l'asma, dermatitis atòpica o rinitis.

L'expressió del fenotip atòpic requereix la interacció d'una predisposició genètica amb al·lèrgens mediambientals. Existeixen diverses regions climàtiques a Europa, i cada una té una distribució d'al·lèrgens diferent. Això implica, que en un estudi multicèntric com el present, s'ha d'escollir una àmplia selecció d'al·lèrgens a provar.

### 2.2.2.1 Elecció de les proves

Poden utilitzar-se dos tipus de proves per determinar l'estatus atòpic, que són:

- a) la mesura de la sensibilitat cutània a al·lèrgens
- b) la mesura de la IgE sèrica com a IgE total o com a IgE específica contra al·lèrgens específics. Requereix extracció de sang.

En Epidemiologia, els estudis de sensibilitat cutània són els més pràctics, ja que són de gran acceptació entre la població general perquè causen molèsties mínimes i no requereixen mostra de sang. Donen també una resposta quantitativa i són relativament barats. Els seus principals inconvenients són que són específics per la exposició a determinats al·lèrgens, sent aquesta exposició molt variable segons el lloc, i que es veuen modificades pels antihistamínics, fenotiazines i imipramines.

La IgE específica és cara d'obtenir, però el sistema *Phadiotop*e dona una resposta absoluta de sí/no sobre si existeix un nivell de IgE significatiu contra algun

dels components d'una bateria d'antígens. Això té alguns inconvenients. És encara específic per un determinat rang d'antígens, requereix una mostra de sang, només dona una resposta qualitativa i és relativament cara. Pel que fa a la IgE total, té baixa sensibilitat i especificitat per atopia, i no es sol usar com a prova clínica.

En el projecte farem servir la IgE a al·lèrgens específics perquè la mesura de sensibilitat cutània resulta difícil d'estandarditzar (Chinn 1996).

### **2.2.2.2 Elecció dels al·lèrgens.**

Existeixen un nombre molt gran d'al·lèrgens. En l'estudi s'ha fet el test de la IgE específica pels al·lèrgens següents:

- *Dermatophagoides pteronyssinus*, àcars de la pols: Els àcars són un grup d'al·lèrgens clínicament molt importants en moltes parts del món. Són aràcnids microscòpics extraordinàriament ubiqües. El creixement d'àcars depèn fonamentalment de dos factors: el grau d'humitat relativa ambiental i la temperatura ambient (la idònia és al voltant d'entre 20 i 25°).
- Herba timotea: l'herba timotea, o *Phleum pratense*, és una gramínia o "espiguilla". Les gramínies es divideixen en dos grans tipus: salvatges i cultivades. Entre aquestes últimes es troben els cereals com el blat, l'ordi o el sègol. No obstant, les més importants des del punt de vista al·lèrgic són les salvatges, que polinitzen a finals de primavera-principis d'estiu, i són predominants de les zones d'interior on, pel seu clima, produeixen símptomes molt intensos encara que curts en el temps. Aquestes plantes són les responsables de la tendència que encara avui en dia existeix a considerar les al·lèrgies un problema exclusivament primaveral. No obstant, a les zones costaneres, on hi ha un predomini d'àcars, l'al·lèrgia sol donar símptomes durant gran part de l'any. Existeix una important reactivitat creuada entre les gramínies,



per la qual cosa normalment s'és al·lèrgic al pol·len de moltes d'aquestes plantes de forma simultània.

- Epteli de gat: un dels al·lèrgens més potents que es coneixen es troba a la saliva del gat, que, degut al seu costum de llepar-se amb freqüència, es diposita als pèls i passa al medi ambient en forma d'un fi halo de partícules. No és necessari el contacte directe amb un gat, ja que tot l'ambient en el que romanen els animals domèstics estarà contaminat. Las escames y el pèl són un constituent important de la pols domèstica i, en pacients que presenten símptomes de forma continuada, és important diferenciar entre els que són deguts als àcars i els que ho són als animals.
- *Cladosporium herbarum*: és un fong, les espores del qual són produïdes amb abundància. Les espores estan presents a l'atmosfera durant tot l'any, amb majors concentracions de maig a octubre, especialment al centre de les ciutats. Colonitza freqüentment fulles i plantes, especialment gramínies. Les espores aerotransportades augmenten notablement quan es sega la gespa o es poden els arbres. Es pot trobar en neveres brutes, en finestres de fusta humida i en cases situades en ambients humits amb escassa ventilació.
- Altres al·lèrgens es van fer servir només en alguns centres i per això no es faran servir en l'anàlisi conjunta. Les anàlisis de laboratori es van fer en un únic laboratori fent servir el mateix sistema i els mateixos instruments.

## 2.3 Metodologia de l'estudi

El protocol de l'ECRHS va ser publicat el 1993 i en una forma abreujada el 1994 (Burney 1994). Els centres participants van seleccionar una àrea definida per fronteres administratives existents amb una població d'almenys 150.000 habitants. Quan va ser

possible, es va fer servir un marc de mostreig actualitzat per seleccionar aleatòriament com a mínim 1.500 homes i 1.500 dones, de la franja dels 20 als 44 anys d'edat.

A la fase I, se'ls va enviar un qüestionari d'*screening* on se'ls demanava per símptomes que podien suggerir asma, l'ús de medicació per l'asma i la presència d'al·lèrgies nasals. La mediana de participació per centres en aquesta fase va ser del 78% , i el rang per anava del 54% al 100%.

A la fase II, una mostra aleatòria més petita de subjectes que havien completat el qüestionari *screening* va ser convidada a participar en la resta de l'estudi, consistent en un qüestionari llarg administrat per un entrevistador, un test cutani (*skin-prick test*), anàlisi de sang per mesurar Immunoglobulina-E específica i total (IgE), espirometria i test de metacolina. La mediana de participació en aquesta fase va ser del 65% (rang 34%-90%). La mostra final conté 18.790 subjectes. En diversos estudis s'han repetit les anàlisis ajustant per la no-resposta usant la informació disponible de la fase I, i aquest ajust pràcticament no ha tingut cap efecte.

L'estudi està centralitzat a Londres i compta amb 5 centres espanyols: Barcelona (centre coordinador a Espanya), Galdakao, Albacete, Oviedo i Huelva.

Els qüestionaris van ser desenvolupats a partir del qüestionari de la *International Union Against Tuberculosis and Lung Diseases* (IUATLD). El qüestionari administrat per un entrevistador contenia, entre altres qüestions, preguntes sobre símptomes respiratoris, asma i al·lèrgia reportada per l'individu, hàbits tabàquics, exposicions ambientals al domicili, ocupació i tractament per l'asma. Les preguntes són les mateixes a cada centre, tot i que hi ha algunes poques preguntes que no van fer-se en alguns centres o d'altres que només es van fer en alguns centres concrets.

El control de qualitat va incloure seminaris d'entrenament abans de l'estudi i visites de control de qualitat a cada centre una vegada durant l'estudi. Un extensiu programa per la qualitat de l'entrada de dades es va posar en pràctica abans d'unir les bases de dades del diferents centres.

Pel que fa a la participació, dels 18.790 individus de la mostra, un 64% accediren a donar mostres de sang. Per tant, es van poder realitzar les proves d'IgE en sang per mesurar l'al·lèrgia a diferents al·lèrgens a 12.033 individus.

## **Capítol 3. Asma i nivell socioeconòmic**

### **3.1. Introducció.**

L'estudi que s'està duent a terme actualment amb aquestes dades a l'IMIM, i que és la base d'aquest projecte, és un estudi del paper de les variables socioeconòmiques en l'asma. Per dur-lo a terme utilitzem dades de tots els centres que disposen de les variables socioeconòmiques que es volen utilitzar. S'estudien tant variables socioeconòmiques a nivell d'individu, com per exemple el nivell màxim d'educació assolit o la classe social a la qual pertanyen els individus (en base a la seva ocupació), com variables a nivell de centre, com per exemple el *Gross National Product* per càpita (GNP per càpita) del país al qual pertany un centre, el percentatge de persones amb un alt nivell d'educació en el centre o el percentatge de persones de classe social baixa del centre.

### **3.2. Antecedents**

L'efecte de variables socioeconòmiques en l'asma s'ha estudiat en alguns estudis, però els resultats no són concloents. Hi ha alguna evidència de relacions entre mortalitat per asma, admissions a l'hospital degudes a l'asma i severitat de l'asma amb pobresa (Volmer 2001, Higgins 1995, Littlejohns 1993, Watson 1996, Walters 1995 i Mielck 1996), però no està clara la influència de la classe social en la prevalença d'asma. Tal i com suggereix Salmond (1999), és convenient distingir entre estudis en nens, que bàsicament tracten tots els nens que en algun moment de la seva vida han tingut asma (incidència acumulada), i estudis en adults, que poden estar influenciats per diferències en la prolongació i l'empitjorament dels símptomes que poden afectar a la durada de la malaltia. No obstant, apareixen resultats contradictoris tant en estudis en nens com en adults.

En adults, alguns estudis que fan servir diverses definicions d'asma i diverses mesures de nivell socioeconòmic basades en ocupació o salari, troben una relació negativa entre asma i classe social alta (Court 2002, Bozicevic 2000, Lewis 2001, Litonjua 1999, Pallasaho 1999, Isoaho 1994, Sibbald 1992 i McWhorter 1989), mentre que Montnemery (2001) troba que no hi ha associació. Quan l'educació és la mesura de nivell socioeconòmic, un estudi troba una relació negativa entre alta educació i asma (Litonjua 1999) i un altre (Chen 2002) no troba associació. En nens, s'han trobat tant relacions positives (Saxena 2002, Nascimento 2002, Peat 1880, Peckham 1978 i Hamman 1975) com negatives (Miller 2001, Alshehri 2000, Schwartz 1990, Weitzman 1990 i Rona 1999) com absència d'associació (Miller 2001, Ernst 1998, SIDRIA 1997, Mitchell 1989 i Gergen 1988).

Alguns autors han suggerit distingir entre asma amb atòpia i asma sense atòpia. Chen (2002) troba una relació positiva entre asma amb rinitis (malaltia similar a l'atòpia) i alt nivell d'educació, mentre que la relació és negativa quan la resposta és asma sense rinitis. Court et al. (2000) també troben una relació negativa entre sibilàncies (xiulets al pit, un símptoma que pateixen els asmàtics) amb presència d'atòpia i classe social alta, però no troba associació quan la resposta és sibilàncies sense atòpia.

A nivell ecològic<sup>2</sup> (a nivell de centre), alguns estudis han examinat la relació entre la prevalença d'asma en diferents àrees i la pobresa de les àrees. Saldmon (1999), Litonjua (1999) i Erzen (1997) troben un increment en la prevalença d'asma en adults en les àrees més empobrides. En nens, Stewart (2001) troba una associació positiva entre la prevalença de sibilàncies en l'últim any i el Gross National Product (GNP) per càpita, mentre que Duran-Tauleria (1999) troba la relació inversa fent servir l'índex de Townsed (un índex de pobresa) per sibilàncies persistents però no troba associació quan la resposta és asma. Burr (1997) tampoc troba associació entre la prevalença d'asma i l'índex de Townsed.

Tant les variables socioeconòmiques individuals com les de nivell de centre poden tenir una influència independent en les malalties. La justificació d'aquest fet està clarament explicada per Díez-Roux (1998). Com ella diu, moltes variables mesurades a nivell individual estan fortament condicionades per processos socials que operen a nivell de grups socials o societats. Alguns sociòlegs han argumentat que la vida dels individus està afectada no només per les seves característiques personals, sinó també per les característiques dels grups socials als quals pertanyen (Blalock 1979, Dogan 1969). La idea subjacent és que les propietats del grup són diferents de les propietats dels membres individuals, i que les variables a nivell de grup poden afectar la resposta independentment de les característiques individuals, o fins i tot modificar la forma amb la qual les característiques individuals estan relacionades amb la resposta (Díez-Roux 1998). Hi ha mesures de pobresa a nivell d'àrea que han estat associades amb variables de salut després d'ajustar pels factors individuals (Morgan 1983, Haan 1987, Hochstim 1968). Per exemple, els ingressos mitjans d'un veïnat poden ser un marcadore d'altres variables a nivell de veïnat relacionades amb la salut (com per exemple els serveis del

---

<sup>2</sup> En Epidemiologia, els estudis en els quals la unitat d'observació són grups d'individus són coneguts com a estudis ecològics

barri pel temps lliure, la qualitat de les escoles, la qualitat de les carreteres o les condicions ambientals), i aquests factors poden afectar a tothom de la comunitat independentment dels seus ingressos individuals (Díez-Roux 1998). Aquests estudis no s'han realitzat fins ara amb l'asma.

### **3.3. Objectiu**

El nostre objectiu és valorar l'associació entre asma i nivell socioeconòmic, tant a nivell individual com a nivell de grup. En l'estudi es fan servir dades de l'*European Community Respiratory Health Survey* (ECRHS) i serà el primer cop que s'avalua l'efecte de variables socioeconòmiques amb aquestes dades de l'ECRHS. Per fer-ho, es tindran en compte totes les variables que en els estudis previs de l'ECRHS han mostrat algun tipus d'associació amb l'asma tant a nivell individual com ecològic.

Aquest estudi és el primer en avaluar el paper en l'asma de les variables socials simultàniament a nivell individual i a nivell de centre. També és el primer estudi que ajusta un model multinivell per intentar explicar l'asma. En el que queda de capítol es justifica l'ús de les variables que s'utilitzaran.

## **3.4 Variables a nivell individual**

### **3.4.1 Variable resposta**

La variable resposta és asma en l'últim any, definida, tal i com s'ha justificat a la secció 2.2.1, com a una o més respostes afirmatives a si en l'últim any s'ha despertat

alguna vegada amb un atac de falta d'aire, o ha tingut algun atac d'asma, o ha pres alguna medicació per l'asma.

### 3.4.2 Variables socioeconòmiques

1. Classe social: la classe social s'ha obtingut a partir de l'ocupació, recollida mitjançant diferents preguntes en el qüestionari que van ser extretes de la *Office of Population Censuses and Surveys* i són compatibles amb el sistema de classificació en grups socioeconòmics de la Unió Europea. A partir d'aquestes preguntes s'ha codificat la classe social fent servir la classificació anglesa *Registrar General Standard Classification of Occupations*, basant-se en l'ocupació actual o, en el cas d'haver canviat la seva ocupació degut als problemes respiratoris que aquesta els provocava, en l'ocupació anterior. Això es fa així per evitar un biaix de selecció. La classificació té 6 categories:

- Classe I: Professionals (ex: metge, advocat, comptable, etc.).
- Classe II: Ocupacions intermèdies (ex: mestre, infermera, etc.).
- Classe III No-manual: treballadors qualificats que treballen en una ocupació no manual (ex: oficinista, secretària, etc.).
- Classe III Manual: treballadors qualificats que treballen en una ocupació manual (ex: carnisser, fuster, etc.)
- Classe IV: Treballadors parcialment qualificats en una ocupació manual (agricultor, carter, conductor d'autobús, etc.)
- Classe V: treballadors no qualificats (ex: netejadors, empleats de fàbrica, etc.)
- No classificables: mestresses de casa, estudiants, militars i gent que no ha contestat les preguntes sobre ocupació.

2. Educació: degut als diferents tipus d'educació existent als països participants, la classificació s'ha fet mitjançant l'any en què van deixar de ser estudiants a temps complet. Els individus s'han dividit en tres categories segons si van deixar d'estudiar abans dels 16 anys, que podríem considerar-se com a gent que ha fet l'escolarització obligatòria, tot i que aquesta edat canvia de país a país i també amb els anys; aquells que van estudiar fins a una edat compresa entre els 16 i els 19 anys, que poden considerar-se com als que han estudiat Batxillerat o Formació Professional; i per últim, aquells que han estudiat fins als 20 anys o més, que podrien considerar-se universitaris.

### 3.4.3 Variables confusores

La majoria de vegades, i sempre en els estudis observacionals, no és suficient amb tenir informació de la variable resposta i del factor d'interès, ja que la relació que hi ha entre els dos pot estar "barrejada" o confosa per l'efecte d'una altra variable, que està correlacionada tant amb la resposta com amb el factor d'interès. Aquestes són les variables confusores, les quals s'han de tenir en compte en l'anàlisi per no atribuir al factor d'interès un efecte que és degut a una tercera variable, o a l'inrevés, per evitar que no trobem efecte del factor d'interès quan en realitat, si es corregeix l'efecte del confusor, sí que té existència aquest efecte. Les possibles variables confusores que es fan servir en el projecte, d'acord amb la literatura sobre el tema, són les següents:

1. Sexe. Se sap que la prevalença d'asma en adults és més alta en les dones (Papageorgiou 1997, Abramson 1996, Plaschke 1999), i poden haver-hi diferències per sexe en el nivell socioeconòmic.
2. Edat: l'estudi estava restringit a edats de 18 a 44 anys. Per tant, l'estudi tracta l'asma en adults joves. La prevalença d'asma autoreportada s'ha mostrat inversament relacionada amb l'edat en alguns estudis (Jarvis 1994, Neukirch 1995, Björnsson 1994, Abramson 1995), i també pot estar relacionada amb el nivell socioeconòmic.



3. Nombre de germans: aquesta variable s'ha mostrat associada amb asma en altres estudis (Sunyer 1997, Jarvis 1997). El mecanisme pel qual aquesta relació apareix no està clar, però s'ha apuntat que pot tenir a veure amb el fet que la gent amb germans estan més exposats a infeccions de petits i el seu sistema immunitari se'n pot veure afectat. També té una relació amb el nivell socioeconòmic.
4. Hàbit tabàquic: al ser l'asma una malaltia respiratòria sembla que l'hàbit pot tenir-hi a veure o actuar com a modificadora. L'associació amb asma, fins ara, no és clara (de Marco 2000, Plaschke 2000, Björnsson 1994, Abramson 1996. També s'han trobat en molts estudis diferències per classe social en el nombre de fumadors. Es tenen en compte els ex-fumadors i la quantitat de cigarrets fumats.
5. Història familiar d'asma: aquesta variable indica si el pare o la mare pateixen o han patit d'asma, i intenta capturar la component genètica i hereditària de la malaltia. L'asma té un control genètic evident (en estudis amb bessons s'atribueix el 50% de l'asma a l'herència i estudis de famílies han identificat marcadors en zones dels cromosomes 5, 11, i altres, que s'associen a diferents registres fenotípics de l'asma com la capacitat de generar anticossos específics a al·lèrgens comuns, la producció d'IgE total o la hiperreactivitat bronquial). Els determinants genètics, no obstant, no poden explicar les variacions geogràfiques en zones genèticament similars. Les interaccions entre els determinants genètics i els factors ambientals seran les que explicaran aquestes variacions. No obstant, història familiar d'asma s'associa fortament amb la prevalença d'asma (Wieringa 1997, Sunyer 1997, Abramson 1996, Chowgule 1998).
6. Test d'IgE de sensibilització a diferents al·lèrgens: Tal i com s'ha explicat a la secció 2.2.2.1, aquesta prova dona un diagnòstic de sí o no respecte a la sensibilització a un al·lèrgè (antigen capaç de desencadenar una reacció al·lèrgica específica d'hipersensibilitat immediata, mitjançant IgE). S'ha mesurat IgE sèrica pels al·lèrgens: *Der p 1* (àcars de la pols), *Fel d 1* (epiteli de gat),

herba timotea, i *Cladosporium Herbarum*. Molts estudis han trobat una relació entre atopia i asma, i s'ha demostrat que la sensibilització a edats joves a al·lèrgens ambientals és un fort predictor de l'aparició posterior a asma (Wieringa 1997, Nowak 1996, Sunyer 1997, Abramson 1996, Plaschke 1999, Chowgule 1998).

7. Infeccions respiratòries en la infància (abans dels 5 anys): aquestes infeccions poden haver afectat el sistema respiratori i el sistema immunitari. Hi ha corrents que pensen que haver tingut infeccions a la infància exerceix un paper iniciador de l'asma, mentre que d'altres asseguren que n'és un inhibidor. La primera possibilitat es basa en que el mal fet a la mucosa bronquial causat pels virus respiratoris podria també facilitar la sensibilització a al·lèrgens inhalats. La segona hipòtesi apunta a que les infeccions víriques durant la primera infància podrien modular el sistema immunitari en el sentit de protegir de l'aparició de l'asma (Björnsson 1996, Ferrari 2000). La relació d'aquesta variable amb el nivell socioeconòmic també té sentit.
8. Rinitis: la rinitis és un tipus d'al·lèrgia nasal que provoca una inflamació de la mucosa del nas, acompanyada d'inflamació i secreció. La rinitis apareix associada amb l'asma en diferents estudis, també en no atòpics (Neukirch 1995, Björnsson 1994, Abramson 1996).
9. Exposicions laborals: es treballa amb la variable alta exposició laboral a pols biològica, pols mineral, gasos o fums. Aquests tipus d'exposicions en l'ambient laboral poden provocar asma, coneguda com a asma ocupacional (Kogevinas 1999). L'exposició laboral s'obté a partir d'una avaluació feta per experts higienistes ocupacionals a partir de les preguntes d'ocupació del qüestionari, que després dissenyen una matriu (Job Exposure Matrix, JEM) que associa a cada ocupació una estimació semi-quantitativa de l'exposició a pols biològica, pols mineral, gasos i fums. Nosaltres treballarem amb una variable dicotòmica que indica una alta exposició laboral a aquests compostos, tal i com suggereix Kogevinas. Aquesta variable, tot i estar inclosa a l'apartat de confusors, pot

també ser considerada com a variable socioeconòmica ja que està relacionada amb l'ocupació.

10. Presència de verdet a la casa: haver tingut verdet a casa durant l'últim any és una pregunta que està relacionada amb l'ambient interior de la casa. La presència de verdet està causada per la humitat interior. El verdet podria provocar asma al·lèrgica a través d'una al·lèrgia als fongs, o també perquè poden provocar inflamació en les vies aèries. No obstant, aquesta variable també pot ser un indicador de classe social. Zock (2002) troba associacions d'aquesta variable amb l'asma.

## **3.5. Variables a nivell de centre o ecològiques**

### **3.5.1. Variables de classe social**

1. GNP per càpita: El *Gross National Product* del país al qual pertany el centre. Aquesta dada ha estat extreta del *World Bank* (1997) i està expressada en dòlars/1000 de 1995.
2. Percentatge d'alta i baixa classe social en el centre: Obtinguda a través d'agregar les variables individuals.
3. Percentatge d'alta i baixa educació en el centre: Obtinguda a través d'agregar les variables individuals.

### **3.5.2. Variables confusores**

1. Percentatge d'història familiar d'asma: obtinguda a través d'agregar les variables individuals. De Vocht (2002) justifica la utilització d'aquesta variable.
2. Temperatura mitjana anual del centre: Verlato (2002) troba una associació ecològica entre la prevalença d'asma dels centres i la temperatura mitjana anual. Aquesta variable s'ha obtingut de la base de dades *Global Historical Climatology Network*, que està produïda conjuntament pel *U.S. Department of Energy's Carbon Dioxide Information Analysis Center* del *Oak Ridge Laboratory* a *Oak Ridge, Tennessee* i pel *National Oceanographic and Atmospheric Administration's National Climatic Data Center* de *Asheville, North Carolina*. També amb es va fer servir les bases de dades AIRBASE, DEM, i altres.
3. Percentatge de fumadors: obtinguda a través d'agregar les variables individuals. Mitchel (2001) troba una relació ecològica entre la prevalença d'asma i el percentatge de fumadors del centre.

# Capítol 4. Descripció dels models i la seva interpretació

## 4. 1. Introducció.

Els models que s'utilitzen en el projecte són models que tenen tant efectes fixos com aleatoris, anomenats per aquesta raó com a models mixtes, encara que també són coneguts com a models multinivell o models jeràrquics, degut a que la població està estructurada en nivells jeràrquics o aniuats (per exemple, en un estudi en diferents empreses els treballadors estan aniuats a l'empresa a la qual pertanyen). Aquests models s'han utilitzat en diferents camps com l'educació, la demografia o la sociologia i permeten examinar simultàniament els efectes de variables a nivell individual i de variables a nivell de grup quan les observacions segueixen una estructura jeràrquica.

Un exemple d'estructura jeràrquica és, per exemple, quan es tenen individus que estan aniuats en famílies, individus que pertanyen a ciutats, o també quan es tenen mesures repetides, on cada individu està mesurat en més d'una ocasió i per tant les ocasions estan aniuades a l'individu. Aquests models no estan limitats només a dos nivells, sinó que poden tenir-ne més. Per exemple, el cas d'individus que pertanyen a barris, a l'hora pertanyen a ciutats, i aquestes ciutats pertanyen a països, correspondria a un model amb quatre nivells. Aquests models permeten la introducció de variables explicatives de qualsevol nivell de la jerarquia. Com a exemple, en el model de quatre nivells comentat anteriorment, podríem incloure tant variables que pertanyin a nivell d'individu com ara l'edat, variables que pertanyen a nivell de barri com el nombre

d'escoles privades que té, variables a nivell de ciutat com el partit polític de l'alcalde que la governa, i variables a nivell de país com per exemple el Producte Interior Brut.

Normalment, quan l'estructura és jeràrquica, les observacions no són independents. Per exemple, els nens d'una escola tendeixen a ser similars degut a un procés de selecció, ja que hi ha escoles que atrauen alumnes d'alt poder socioeconòmic i d'altres de baix. Així doncs, la correlació entre dues variables mesurada en individus de la mateixa escola serà més alta que si la mesurem en individus de diferents escoles. Els tests estadístics estàndard es basen en la hipòtesi d'independència de les observacions, i si aquesta hipòtesi és violada, com passa en un cas com el descrit, els errors estàndard obtinguts amb aquests mètodes són massa petits, produint més resultats significatius del compte. Els models multinivell ja tenen en compte aquesta no-independència de les observacions.

En els últims anys, l'interès per aquests models en la salut pública ha crescut molt. Aquest creixement ha estat estimulat en part pel ressorgiment de l'interès en els potencials determinants ecològics de la salut i per la noció que variables a nivell de grup poden ser rellevants per entendre la distribució de diferents malalties. Una segona raó per l'augment de l'ús d'aquests mètodes és l'acceleració en el desenvolupament de tècniques estadístiques i de software, i el fet que aquests models tinguin aplicacions en un ampli ventall de circumstàncies on l'estructura de la població és aniuada.

Els avenços en aquests tipus de models, particularment en els models lineals generalitzats, i sobretot les implementacions en software, s'han fet recentment. Bàsicament la història d'aquests models va lligada a trobar mètodes per poder-los estimar. A la secció 4.5 s'expliquen els mètodes existents i es pot veure l'evolució històrica d'aquest tipus de models i els tipus d'estimació que s'utilitzen. Bàsicament, les primeres referències les trobem l'any 1984 i 1985, quan Stirateli, Laird i Ware (1984) per una banda i Wong i Mason (1985) per una altra van descriure dos tipus de mètodes d'aproximació pel model jeràrquic de regressió logística. Pel que fa al software, en S-Plus encara no hi ha una funció específica per aquests models, tot i que disposa d'una versió beta desenvolupada per José Pinheiro. En SAS existeix el PROC GLIMMIX,

però fa servir un mètode aproximat que pot no ser bo per fer inferència en alguns casos. Actualment tenen una versió experimental que ho millora, el PROC NLMIXED. Stata té una comanda per aquests models però només funciona a una velocitat adequada a la versió 8. També s'han desenvolupat alguns paquets específics per models multinivell, com per exemple MLwiN, HLM i MIXREG/MIXOR.

Els models d'efectes aleatoris en models lineals generalitzats, com els que es fan servir en aquest projecte, no s'estudien a la Llicenciatura en Ciències i Tècniques Estadístiques, i encara costen de trobar al software i als llibres, on per exemple, no tots els llibres de models multinivell parlen de models lineals generalitzats, i els que ho fan hi dediquen només un capítol.

En les següents seccions es descriuen estadísticament aquest tipus de models. Com a referències he fet servir els textos de Snijders and Bosker (1999), Wong and Mason (1985), Goldstein (1995), Hox (1995), Breslow and Clayton (1993) i Diez-Roux (2000).

## 4.2. Quan utilitzar efectes aleatoris

S'utilitzen efectes aleatoris, en lloc del tractament habitual amb efectes fixos, quan els grups que tenim no són d'interès individualment, sinó que representen una font de variació en la qual estem interessats. És a dir, no es volen conèixer les diferències que existeixen entre el grup 1 i el 2, o entre el 4 i el 7, sinó que el què es vol és una estimació de la variabilitat que existeix entre tots els grups de la població, i per obtenir-la, d'entre tots els possibles grups, se n'escullen uns quants com a representants. El que es vol conèixer és la relació entre la variació entre grups i la variació individual o intra-grups. Per exemple, en el cas que es tingui un sol factor aleatori, s'està suposant que els  $K$  grups són una mostra aleatòria de la població de grups. Encara que no sempre és veritat que els grups s'hagin escollit aleatòriament, sinó que moltes vegades s'escullen

per raons de cost o de disponibilitat, aquesta hipòtesi serveix com a indicació que els grups són d'interès només com a representants d'un tipus de variació.

Això es pot il·lustrar en un model ANOVA d'un sol factor. El model d'efectes fixos suposa que  $y_{ij} \sim N(\mu + \tau_i, \sigma_e^2)$  independents, i la prova d'hipòtesi més natural és

$$\begin{cases} H_0 : \tau_i = 0 \quad \forall i \\ H_a : \exists i \text{ tq } \tau_i \neq 0 \end{cases},$$

mentre que el model d'efectes aleatoris suposa que

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

on  $\tau_i \sim N(0, \sigma_\tau^2)$  i  $\varepsilon_i \sim N(0, \sigma_e^2)$  són independents entre ells, i per tant que

$$y_{ij} \sim N(\mu, \sigma_\tau^2 + \sigma_e^2), \text{ de manera que } \text{cov}(y_{ij}, y_{kr}) \begin{cases} 0 & i \neq k \\ \sigma_\tau^2 & i = k \end{cases}.$$

La seva prova d'hipòtesi més natural és:

$$\begin{cases} H_0 : \sigma_\tau = 0 \\ H_a : \sigma_\tau \neq 0. \end{cases}$$

Com es pot veure, en el model d'efectes fixos es fa inferència sobre els paràmetres que modelen el valor esperat de  $y_{ij}$ , mentre que en el d'efectes aleatoris es fa inferència sobre el paràmetre que regula la matriu  $\Sigma$  de variàncies i covariàncies. Per tant, la forma d'analitzar-ho és diferent.

Les conclusions a les quals s'arriba en cada una de les dues aproximacions també són diferents. En cas de rebutjar la hipòtesi nul·la en efectes fixos, es conclou que els grups de la mostra no són tots iguals, i es pot prosseguir fent comparacions múltiples per saber quin és diferent de quin. En el cas d'efectes aleatoris això no és el que interessa. En efectes aleatoris, quan rebutgem la hipòtesi nul·la conclouem que existeixen diferències entre grups en tota la població de grups, tant els que estan a la mostra com els que no hi són. Haver inclòs una component aleatòria a nivell de grup és el que permet concloure per tota la població de grups (si la mostra de grups escollida és bona).



## 4.3. Models lineals mixtes

Els models que faig servir en el projecte no són models lineals mixtes, sinó que són models lineals generalitzats mixtes. No obstant, els explico perquè són més fàcils d'entendre i un cop entesos, el pas per entendre els models lineals generalitzats és relativament senzill.

### 4.3.1. Model amb la constant aleatòria

Considerem un model de dos nivells, en el qual tenim  $J$  unitats del nivell més baix (nivell 1) que pertanyen a  $I$  grups (nivell 2), on el grup  $i$ -èssim té  $n_i$  unitats de nivell 1. El model de regressió lineal simple amb efectes aleatoris més senzill que es pot aplicar és el que considera la constant del model com a efecte aleatori. Aquest model ajusta per cada grup una constant diferent, però en canvi el pendent, que en aquest cas és un efecte fix, és el mateix per tots els països. L'expressió del model és la següent:

$$y_{ij} = \beta_{0i} + \beta_1 x_{ij} + \varepsilon_{ij}, \quad \text{on } i = 1..I, \quad j = 1..n_i. \quad (1)$$

$\beta_{0i}$  es pot expressar com un valor constant (igual per tots els grups) més una perturbació, o residu a nivell de grup,  $u_i$ , que assumim independent de  $\varepsilon_{ij}$ :  $\beta_{0i} = \gamma_0 + u_i$ , on,  $u_i \sim N(0, \sigma_u^2)$ . Així doncs, el model complet és el següent:

$$\boxed{\begin{array}{l} y_{ij} = \gamma_0 + \beta_1 x_{ij} + u_i + \varepsilon_{ij} \quad \text{on } i = 1..I, \quad j = 1..n_i, \\ u_i \sim N(0, \sigma_u^2), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \quad \text{independents} \end{array}}, \quad (2)$$

o, de forma compacta,  $y_{ij} \sim N(\gamma_0 + \beta_1 x_{ij}, \sigma_u^2 + \sigma_\varepsilon^2)$  amb

$\text{Cov}(y_{ij}, y_{kr} | x_{ij}, x_{kr}) = \begin{cases} 0 & i \neq k \\ \sigma_u^2 & i = k \end{cases}$ , on  $\gamma_0$ ,  $\beta_1$  i  $\sigma_u^2$  són les incògnites que es volen estimar.

Els models que només tenen efectes aleatoris a la constant, són coneguts com a models de components de la variància, ja que la variància total queda dividida en la suma de la variància deguda al grup més la variància residual. En aquests cas, la correlació entre dos individus del mateix grup serà  $\text{Corr}(y_{ij}, y_{ik} | x_{ij}, x_{ik}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$ , i és el que en alguns entorns anomenen correlació intra-grup. Analitzant-la s'observa que és el percentatge de la variabilitat total de la resposta que és deguda a la variabilitat entre grups.

Si es vol saber si la constant del model és diferent entre grups, cal fer el test d'hipòtesi  $H_0: \sigma_u^2 = 0$ . Si rebutgem la hipòtesi nul·la, existeixen diferències estadísticament significatives entre grups en les constant  $\beta_{0i}$ . En aquest cas, s'obté una recta diferent per cada grup, però com que el que varia només és la constant, les rectes seran paral·leles.

### 4.3.2. Models amb pendents aleatoris

A partir del model de l'apartat anterior, es pot considerar també que, a part de que la constant sigui diferent en cada grup, el pendent de la variable explicativa també pugui variar. Així, en cada grup tenim un model diferent, tant en  $\beta_0$  com en  $\beta_1$ . L'expressió del model és la següent:

$$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \varepsilon_{ij} \quad \text{on } i = 1..I, \quad j = 1..n_i. \quad (3)$$

Assumint novament que els coeficients aleatoris es distribueixen aleatòriament al voltant d'una constant,  $\beta_{0i} = \gamma_0 + u_{0i}$  i  $\beta_{1i} = \gamma_1 + u_{1i}$ , on  $u_{0i} \sim N(0, \sigma_{u0}^2)$ ,  $u_{1i} \sim N(0, \sigma_{u1}^2)$  i  $\varepsilon_{ij} \sim N(0, \sigma_e^2)$ . Substituint les últimes expressions a l'expressió general s'obté:

$$y_{ij} = \gamma_0 + u_{0i} + (\gamma_1 + u_{1i}) x_{ij} + \varepsilon_{ij} = \gamma_0 + \gamma_1 x_{ij} + u_{0i} + u_{1i} x_{ij} + \varepsilon_{ij}. \quad (4)$$

S'observa que el nou efecte aleatori,  $u_{1i}$ , apareix multiplicant a la variable explicativa  $x$ . A més a més, els dos efectes aleatoris  $u_{0i}$  i  $u_{1i}$  són del mateix nivell, i per

tant pot existir correlació entre ells. Aquest nou terme de correlació el notarem  $u_{01}$ . Tot això tindrà conseqüències a l'hora de calcular la variància i la covariància, ja que:

$$E(y_{ij}|x_{ij}) = \gamma_0 + \gamma_1 x_{ij},$$

$$V(y_{ij} | x_{ij}) = \sigma_{u0}^2 + \sigma_{u1}^2 x_{ij}^2 + 2\sigma_{u01}x_{ij} + \sigma_e^2,$$

$$\text{Cov}(y_{ij}, y_{ir} | x_{ij}, x_{ir}) = \sigma_{u0}^2 + \sigma_{u1}^2 x_{ij} x_{ir} + \sigma_{u01}(x_{ij} + x_{ir}).$$

Ara, la variància deixa de ser constant ja que depèn de la  $x$ , i per tant no es verifica ni la hipòtesi d'homocedasticitat ni tampoc la d'independència, ja que dos individus d'un mateix grup tenen una correlació diferent de zero. També cal observar que l'estructura de variàncies i covariàncies es complica molt cada vegada que afegim un efecte aleatori nou. En aquest cas, les rectes per cada grup deixen de ser rectes paral·leles ja que el pendent també varia de grup a grup.

El model es pot estendre a un model amb  $P$  variables explicatives. L'expressió, amb el subíndex  $p$ , on  $p=1, \dots, P$ , indicant les diferents variables explicatives, és la següent:

$$Y_{ij} = \gamma_0 + \sum_{p=1}^P (\gamma_p + u_{pi})' X_{pij} + u_{0i} + \varepsilon_{ij}, \quad (5)$$

on  $u_{pi}$  és zero si la variable  $p$ -èssima no té el pendent aleatori.

### 4.3.3 Introducció de variables de nivell superior

Com ja s'ha dit, els models multinivell permeten que el model tingui variables explicatives a qualsevol nivell de la jerarquia. Aquest subcapítol amplia el model de l'apartat anterior mitjançant la inclusió de variables de nivell superior.

El fet de tenir pertorbacions a més d'un nivell, ja que tenim les pertorbacions  $\varepsilon_{ij}$ , de les quals n'hi ha una per cada individu, i les pertorbacions  $u_{pi}$ , que en tenim una per cada grup per la  $p$ -èssima variable, permet que puguem introduir variables a nivell de grup al model (variables amb valor constant dins d'un grup). Per exemple, variables

ecològiques com la riquesa del país o el nombre d'habitants. Aquestes variables poden servir per explicar part de la variància entre grups.

Sigui  $Z_i$  una variable a nivell de grup. Aquestes variables es poden entrar al model de diverses maneres:

- a) La manera més senzilla és suposar que la constant de cada grup està explicada per la variable a nivell de grup. Partint del model de l'equació (3), i suposant que :

$$\begin{aligned}\beta_{0i} &= \gamma_{00} + \gamma_{01} Z_i + u_{0i} \quad , \\ \beta_{1i} &= \gamma_{10} + u_{1i} \quad ,\end{aligned}\tag{6}$$

substituint i desenvolupant el model obtenim:

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} Z_i + u_{0i} + u_{1i} x_{ij} + \varepsilon_{ij} \quad .\tag{7}$$

Si ens fixem en l'expressió final, tenim un paràmetre que actua com a constant del model,  $\gamma_{00}$ , un pendent que acompanya la variable explicativa  $x$ ,  $\gamma_{10}$ , i ara a més a més tenim un altre pendent per la variable a nivell de grup  $z$ ,  $\gamma_{01}$ . Aquest darrer pendent, s'interpreta de la mateixa manera que els altres. És a dir, per cada unitat que augmenta la variable  $z$ , si totes les altres variables es mantenen fixes, la resposta  $y$  augmenta en  $\gamma_{01}$ . També pot interpretar-se a partir de l'equació (6), és a dir, per cada unitat que augmenta la  $z$ , la constant  $\beta_{0i}$  del grup  $i$  augmenta en  $\gamma_{01}$ .

Cal observar que estem assumint que una variable a nivell de grup pot influir en la variable resposta mesurada a nivell individual. Per exemple, en una regressió entre els resultats acadèmics d'alumnes de diferents escoles i el nivell socioeconòmic, el fet d'estar en una escola d'un barri amb alt nivell socioeconòmic pot fer que les puntuacions d'un alumne siguin més altes, independentment del nivell socioeconòmic individual de l'alumne. Així doncs, amb aquest tipus de models podem corregir l'efecte confusor de variables a nivell de grup. De fet, obtenim l'efecte de les variables individuals i de grup ajustats mútuament, ja que totes estan incloses en el mateix model.

b) Una segona manera d'introduir una variable a nivell de grup al model és permetent que, a part de que la variable a nivell de grup expliqui per què una constant és més alta en un grup que en un altre, expliqui també les variacions entre grups en el pendent. És a dir, que ara tant la constant com el pendent estan explicats per la variable a nivell de grup. Partint novament de l'equació (3), a més a més de (6) també suposem que:

$$\beta_{1i} = \gamma_{10} + \gamma_{11} Z_i + u_{1i} \quad , \quad (8)$$

i substituint-ho a l'expressió general, obtenim:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_i + \gamma_{11} Z_i * X_{ij} + u_{0i} + u_{1i} X_{ij} + \varepsilon_{ij} \quad . \quad (9)$$

El nou paràmetre que apareix en aquest model,  $\gamma_{11}$ , és un terme d'interacció entre variables de diferents nivells, i per això a aquests termes se'ls coneix com a “cross-level interactions”. Si aquest terme d'interacció és estadísticament significatiu, la variable a nivell de grup modifica l'efecte de la variable individual.

Fent servir una relació com la de l'equació (8) podem fer que els pendents d'altres variables també estiguin explicats per variables a nivell de grup. El model multinivell complet, incloent  $P$  variables explicatives individuals  $X_p$ , amb  $p=1, \dots, P$ , i  $Q$  variables explicatives a nivell de grup  $Z_q$ , amb  $q=1, \dots, Q$ , i amb les interaccions que siguin necessàries té la següent notació:

$$Y_{ij} = \gamma_{00} + \sum_{p=1}^P (\gamma_{p0} + u_{pi}) X_{pij} + \sum_{q=1}^Q \gamma_{0q} Z_{qi} + \sum_{p=1}^P \sum_{q=1}^Q \gamma_{pq} Z_{qi} X_{pij} + u_{0i} + \varepsilon_{ij} \quad (10)$$

### 4.3. Models logístics mixtes

L'extensió dels models de l'apartat anterior a models lineals generalitzats també és possible. En particular, en aquest projecte final de carrera es fan servir models de regressió logística amb efectes aleatoris.

La resposta observada  $y_{ij}$  en aquests models són proporcions amb l'assumpció que, condicional a conèixer  $x_{ij}$ , segueixen la distribució binomial,  $y_{ij} \sim \text{Bin}(\pi_{ij}, n_{ij})$ , on  $n_{ij}$  és el denominador de la proporció. Per modelar-ho farem servir la funció de link logit, que és la més usual i permet expressar els resultats en odds ratios, que tenen una interpretació més clara que els resultats que es puguin obtenir, per exemple, amb el link probit.

Considerem un model de components de la variància amb dos nivells, amb una proporció com a resposta i amb el link logit. L'expressió, en funció de la probabilitat és la següent:

$$\pi_{ij} = \{1 + \exp(-[\beta_0 + \beta_1 x_{1ij} + u_{0i}])\}^{-1}, \quad (11)$$

on  $u_{0i} \sim N(0, \sigma_{u_0}^2)$ ,

expressió que és lineal respecte al logaritme dels odds:

$$\log(\pi_{ij} / (1 - \pi_{ij})) = \beta_0 + \beta_1 x_{1ij} + u_{0i}.$$

A partir de l'última expressió és fàcil, seguint les passos de les seccions 4.3.1-4.3.3, generalitzar-ho a models amb més d'una  $x$ , amb pendents aleatoris, amb variables a nivell de grup i amb interaccions entre variables de diferents nivells, per acabar tenint l'expressió general:

$$\log\{\pi_{ij} / (1 - \pi_{ij})\} = \gamma_{00} + \sum_{p=1}^P (\gamma_{p0} + u_{pi})' X_{pij} + \sum_{q=1}^Q \gamma_{0q}' Z_{qi} + \sum_{p=1}^P \sum_{q=1}^Q \gamma_{pq} Z_{qi} X_{pij} + u_{0i}$$

## 4.4. Interpretació dels efectes aleatoris.

Els termes  $u_{0i}$  (equació 10) mesuren la desviació de la constant de cada centre de la constant global  $\gamma_{00}$ , després d'ajustar per la resta de variables incloses en el model. Aquests "errors" a nivell 2 segueixen una normal, amb mitjana 0 i variància  $\sigma_{u_0}^2$ . Anàlogament, els termes  $u_{pi}$ , amb  $p=1, \dots, P$ , són les desviacions del pendent de la p-èsima variable de cada grup del pendent de la p-èsima variable global, després de d'ajustar per la resta de variables. Pel que fa als "errors" d'aquests pendents també

s'assumeix que es distribueixen seguint una normal, amb mitjana zero i variància  $\sigma_{up}^2$ . Així, els models multinivell resumeixen la distribució dels coeficients específics de cada grup mitjançant dues parts: una de fixa que no canvia entre grups,  $(\gamma_{00}, \gamma_{p0})$ , i una d'aleatòria que pot variar de grup a grup,  $(u_{0i}, u_{pi})$ . La hipòtesi que hi ha al darrere, des d'un punt de vista freqüentista, és que les constants i els coeficients de cada grup són mostres aleatòries d'una població normal de constants i coeficients de tots els grups que existeixen.

L'estimació de  $\sigma_{u0}^2$  i  $\sigma_{up}^2$  i el veure com canvien quan s'introdueixen al model variables a nivell d'individu i a nivell de grup permet una quantificació de la variabilitat entre grups i del grau amb el qual aquesta està estadísticament explicada per les característiques dels individus i dels grups. Així, els models multinivell es poden usar per examinar si els efectes de les variables a nivell d'individu varien entre grups o si aquesta variació està explicada per variables a nivell de grup. Els models multinivell permeten la quantificació de la variació a diferents nivells: dins dels grups, quantificada per  $\sigma_e^2$ , i entre grups, quantificada pels efectes aleatoris  $\sigma_{u0}^2$  i  $\sigma_{up}^2$ .

Així doncs, a la pràctica, si  $\sigma_{u0}^2$  és significativament diferent de zero implica que  $\beta_0$  (la constant del model) és estadísticament diferent entre grups (és heterogènia entre grups). Si  $\sigma_{up}^2$  és diferent de zero, el pendent (i per tant l'efecte) de la variable p-èsima és estadísticament diferent entre grups. En el cas del model logístic, el fet que el pendent de la p-èsima variable sigui heterogeni entre grups implica per extensió que l'odds ratio d'aquella variable és heterogeni entre grups. En el cas de la constant,  $\exp(\beta_0)$  és en el model logístic l'odds de la resposta per la categoria de referència, és a dir, quan la resta de variables explicatives del model són zero. Per tant, si detectem heterogeneïtat en  $\beta_0$ , l'odds de la resposta a la categoria basal és heterogeni entre grups.

## 4.5. Mètodes d'estimació i software

En un model multinivell de regressió lineal, si coneguéssim els valors de les variàncies, es podria aplicar directament mínims quadrats generalitzats (GLS) per obtenir estimacions dels coeficients fixos. Quan els residus segueixen una distribució Normal, aquests estimadors són també els màxims versemblants. Al no conèixer les variàncies, es fa servir un procediment iteratiu: mínims quadrats generalitzats iteratius (IGLS). El procediment de màxima versemblança produeix estimadors esbiaixats dels paràmetres aleatoris perquè no té en compte la variació mostral dels paràmetres fixos. Aquest biaix pot ser important en mostres petites. Podem produir estimadors no esbiaixats fent servir una modificació coneguda com a màxima versemblança restringida (REML). L'algoritme IGLS es pot modificar per produir estimadors restringits (RIGLS).

L'estimació dels paràmetres en models lineals generalitzats jeràrquics és més complicada que en els models lineals jeràrquics. Inevitablement s'han de fer servir algun tipus d'aproximació, i se n'han proposat diverses. Els mètodes que s'utilitzen més freqüentment estan basats en la descomposició de la funció de link en funcions de primer o segon ordre segons la fórmula de Taylor. Quan l'aproximació es fa l'entorn de l'estimació de la part fixa s'anomena quasi-likelihood marginal (MQL), i quan és al voltant d'una estimació de la part fixa més la part aleatòria s'anomena quasi-likelihood penalitzada o predictiva (PQL) (Breslow i Clayton 1993, Goldstein 1991). Aquests procediments estan implementats en el paquet *MLwiN*, i també en el PROC *GLIMMIX* de SAS. Una aproximació per Laplace, que se suposa que és més precisa, està implementada en la versió 5 d'un altre paquet anomenat *HLM*. Hi ha altres mètodes d'aproximació com els que proposa Wong i Mason (1985) i Engel i Keen (1994).

Altres procediments fan servir integració numèrica, com els proposats per Stiratelli, Laird i Ware (1984), Anderson i Aitkin (1985), Gibbons i Bock (1987), i Longford (1994). Gibbons i Hedeker (1994) mostren un exemple, que es va estendre a models de tres nivells el 1997 (Gibbons i Hedeker, 1997). El mètode d'integració



numèrica està implementat en el programa MIXOR (Hedeker i Gibbons 1996) i també en la comanda gllamm d'Stata. Una avantatge pràctica d'aquest mètode és que produeix una deviança que es pot fer servir per fer tests d'hipòtesi. La deviança que produeixen els mètodes MQL i PQL, en canvi, no es pot fer servir per això. Cal notar que els paràmetres aleatoris que estima MIXOR són les desviacions estàndard  $\sigma_0$ , etc. en lloc de les variàncies  $\sigma_0^2$ , i hi ha fórmules que descriuen com passar dels errors estàndard de  $\sigma_0$  als de  $\sigma_0^2$ .

També s'han proposat mètodes de càlcul intensiu relacionats amb el bootstrap i el Gibbs sampling (Zeger i Karim 1991, Kuk 1995, Meijer et al. 1995 i McCulloch 1997). Alguns procediments estan implementats en MlwiN. Un mètode basat en el principi d'inferència indirecta va ser proposat per Mealli i Rampichini (1999). Un altre mètode de càlcul intensiu és el mètode dels moments simulats. Aquest mètode s'aplica als models multinivell per models lineals generalitzats a Gouriéroux i Montfort (1996) i una revisió dels treballs recents es troba a Baltagi (1995).

Els estimadors obtinguts per cada un d'aquests mètodes difereixen bàsicament respecte als paràmetres de la part aleatòria. Els estimadors per la part fixa normalment no són gaire diferents. Però especialment si els components de la variància són bastant grans, aquests mètodes poden produir estimacions bastant diferents pels paràmetres de la part aleatòria, que afecten llavors als errors estàndard dels paràmetres fixos.

Els estimadors dels paràmetres aleatoris quan fem servir l'aproximació de primer ordre pels mètodes MQL i PQL tenen un biaix que fa que les estimacions siguin més baixes del que han de ser (Rodríguez i Goldman, 1995). Les estimacions de segon ordre produeixen estimacions amb menys biaix, però pel que sembla, amb un error quadràtic mig més gran.

La integració numèrica per Laplace sembla produir estimacions estadísticament més satisfactòries que no els mètodes MQL i PQL. A més, té l'avantatge de produir una deviança útil pels tests, mentre que la que produeixen MQL i PQL no ho és. El que passa és que aquest mètode està implementat en pocs paquets.

El procés d'estimació d'aquests models és encara una àrea de recerca activa. Triar entre aquests mètodes s'ha de fer en base a l'estabilitat de l'algoritme, és a dir, si convergeixen a una estimació vàlida, l'eficiència estadística, la disponibilitat de software, i la possibilitat d'obtenir tests pels paràmetres.

Els algoritmes que estan disponibles actualment no són perfectament estables; convergeixen segons el conjunt de dades, la complexitat del model que s'ajusta i els valors inicials. Grandàries de grup petites poden contribuir a la inestabilitat de l'algoritme. Dels mètodes MQL i PQL de primer i segon ordre, el mètode MQL de primer ordre és el més estable.

En el meu cas he fet servir el paquet estadístic MlwiN, que és el que tenia disponible per fer-ho. Els models els he estimat pel mètode PQL de segon ordre, encara que provant el mètode MQL de primer o segon ordre les diferències són mínimes.

Els testos de la part fixa del model es poden fer amb t-tests o Wald test. No obstant, els testos pels paràmetres aleatoris són més complicats que en un model jeràrquic lineal. Si es fa servir un programa que implementi integració numèrica, com per exemple MIXOR, o l'aproximació de Laplace, que fa servir la versió 5 de HLM, la deviança es pot fer servir per produir un test de chi-quadrat, però si es fa servir els mètodes MQL o PQL s'obtenen només aproximacions que no són fiables per usar en testos de deviances.

Per aquesta raó he repetit els models amb el paquet MIXOR, que es pot baixar gratuïtament de la web (<http://tigger.uic.edu/%7Ehedeker/mix.html>), sobretot de cara a fer els testos per la part aleatòria, ja que no és recomanable utilitzar els mètodes MQL i PQL per fer inferència, i només es recomanen per fer exploració dels models. Els resultats obtinguts amb els dos paquets varien poc pel que fa als paràmetres fixos i també pels aleatoris, encara que els errors estàndards dels paràmetres fixos són en general bastant més alts que els obtinguts amb MlwiN. No obstant això, les conclusions bàsiques del projecte no es veuen afectades.

# **Capítol 5. Estratègies de modelització, validació i interpretació**

## **5.1. Estratègia d'anàlisi**

### **5.1.1. Anàlisi exploratòria de dades**

Degut a que estem interessats en la variabilitat entre centres, la descripció de les dades es farà estratificada per centre, per així poder valorar les diferències que existeixen entre centres i també per poder caracteritzar-los.

L'anàlisi estarà estructurada tal i com ja s'ha estructurat el capítol 3 en la descripció de les variables, és a dir, mantenint una separació entre variables a nivell d'individu i a nivell de centre, i dins d'aquests dos grups, començar amb les variables d'interès, que són les variables socioeconòmiques, i després fixar-nos en les possibles variables confusores.

Així, començaré mostrant la participació i el nombre final d'individus de cada centre. Seguidament, descriuré la prevalença d'asma en cada un dels centres. Les següents taules descriuran els percentatges o mitjanes de la resta de variables, primer de les variables socioeconòmiques i després de les confusores, també per cada centre.

Posteriorment em fixaré en com es relacionen les tres variables socioeconòmiques (classe social, educació i exposició laboral a contaminants) entre elles, per avaluar fins a quin punt aquestes variables estan explicant la mateixa cosa.

El següent pas és fixar-nos en les variables a nivell de centre, i tal i com s'ha fet amb les variables individuals, primer ho faré només a nivell descriptiu, per poder-nos fer una idea de quins valors prenen les variables en la població que estem estudiant.

Un cop descrites totes les variables, començaré a creuar-les amb la variable resposta asma, seguint l'ordre habitual: primer ens fixarem en la prevalença d'asma en cada una de les categories de les variables socioeconòmiques, i posteriorment descriurem aquesta prevalença en les categories de les variables confusores.

Com que les característiques que ha de complir una variable confusora són que 1) ha d'estar relacionada amb la resposta i 2) ha d'estar relacionada amb la variable explicativa que estem estudiant, creuarem les variables confusores amb les variables socioeconòmiques. Això ens permetrà avaluar quines variables poden confondre la relació i a entendre per què ho fan.

Per examinar com les variables a nivell de centre poden influenciar l'asma, mostraré gràfics de la prevalença d'asma en cada centre contra les variables a nivell de centre. Per avaluar numèricament com es correlacionen aquestes variables utilitzaré el coeficient de correlació d'Spearman.

La correlació d'Spearman està basada en els rangs de les dades, és a dir, en les posicions relatives de cada valor. A l'igual que la correlació de Pearson produeix valors entre  $-1$  i  $1$  amb exactament la mateixa interpretació. La correlació d'Spearman és un mètode no paramètric i per tant no necessita que les dades estiguin distribuïdes de cap manera. Una de les assumpcions que s'ha de fer perquè la correlació de Pearson sigui vàlida és que la distribució conjunta de les dues variables  $X$  i  $Y$  que es volen correlacionar sigui (aproximadament) gaussiana. En mostres grans aquesta assumpció no és tant important, però en el nostre cas tenim només 32 punts, ja que tenim 32

centres participants. A més a més, moltes de les nostres variables a nivell de centre presenten asimetria. Si no es vol assumir que la distribució és gaussiana una bona opció és fer servir el coeficient de correlació d'Spearman, que és no paramètric i per tant no fa aquesta assumptió.

Per calcular-lo es passa de les dades de les variables X i Y als rangs d'ordre. És a dir, cada valor de X es substitueix pel rang que té: la observació més alta es substitueix per un 1, la segona més alta per un 2, i així successivament fins a la N-èsima, que és la observació més baixa de la variable. Després es fa el mateix amb la variable Y i es calcula la suma de quadrats de les diferències (D) entre els rangs de X i els de Y. La fórmula final, on N és el nombre d'observacions, és la següent:

$$r_s = 1 - \frac{6 \sum D^2}{N(N^2-1)}$$

En la següent taula es pot veure un petit exemple on la correlació d'Spearman és  $r_s=0.83$ :

Id	X	Y	Rang X	Rang Y	D=Rang X-Rang Y	D <sup>2</sup>
1	100	145	1	2	-1	1
2	88	150	2	1	1	1
3	77	110	3	5	-2	4
4	56	132	4	3	1	1
5	43	111	5	4	1	1
6	32	100	6	7	-1	1
7	21	91	7	8	-1	1
8	15	87	8	6	2	4
N=8						$\sum D^2 = 14$

Per qualsevol conjunt de parelles de rangs, el mínim de  $\sum D^2$  es dona quan els rangs de les dues variables coincideixen i és zero. El màxim es dona quan hi ha una correlació negativa perfecta, i en aquest cas  $\sum D^2 = N*(N^2-1)/3$ . Llavors, el quocient de  $\sum D^2$  amb el seu màxim és  $3*\sum D^2 / N*(N^2-1)$ , i és zero quan la correlació és perfecta i positiva, 1 quan és perfecta però negativa, i 0.5 quan no hi ha correlació. Perquè prengui els valors habituals d'una correlació (-1 = correlació negativa, 0 = no correlació, 1 =

correlació positiva), cal multiplicar per 2 l'últim quocient i restar-lo a 1, obtenint la fórmula final de la correlació d'Spearman prèviament anunciada.

## **5.1.2. Modelització**

### **5.1.2.1. Estratègia escollida**

L'especificació del model en un model jeràrquic consisteix en escollir, per la part fixa, quines són les variables explicatives rellevants i les seves interaccions, i per la part aleatòria, escollir quins són els pendents aleatoris. L'especificació del model és una de les parts més difícils de la inferència estadística, perquè hi ha dos parts que hi intervenen: una que depèn de la persona que ajusta el model i una altra que es basa en consideracions estadístiques. L'objectiu és arribar a un model que descrigui les dades observades fins a un punt satisfactori però sense complicacions innecessàries. Un propòsit paral·lel és obtenir un model que estigui basat en la part de les dades que té un interès substancial i no en la part que és deguda a l'atzar.

L'estratègia a seguir a l'hora de construir un model de regressió lineal ja és complicada, però en un model multinivell el nombre de complicacions es multiplica degut a la complicació que té la part aleatòria. En comparació amb un model d'un nivell, ara cal modelar a més a més efectes aleatoris per la constant i pels pendents, tenim variables de més d'un nivell, i tenim possibles interaccions entre variables de nivells diferents.

El fet que els models siguin ja de per si complicats, combinat amb el grau de subjectivitat que hi pot haver a l'hora de construir el model, fa que no hi hagi unes regles fixes a seguir. L'especificació del model ha de ser un procés guiat pels següents principis, segons Snijders & Bosker (1999):

- Consideracions relacionades amb el tema que s'estudia, que provenen del camp d'estudi, de la teoria existent, i també del sentit comú.
- La distinció entre efectes que a priori designem com als efectes d'interès, que són en els que es centra l'estudi, i els altres efectes, que són necessaris per obtenir un bon ajust del model. Sovint, els efectes d'interès són una part dels efectes fixos del model, i la part aleatòria té un interès secundari. Quan no hi ha un fort coneixement previ sobre quines variables poden tenir una part aleatòria, es pot seguir una aproximació basada en el que ens diguin les dades per escollir quina és la part aleatòria del model.
- Si un model conté interaccions, els efectes principals han de romandre en el model encara que no siguin significatius per evitar interpretacions errònies. I si una variable té un pendent aleatori, normalment s'ha d'incloure encara que l'efecte fix no sigui significatiu.
- Quan hi ha un efecte aleatori important en la constant del model, hi ha variacions no explicades importants entre les mitjanes dels diferents grups. Es pot mirar de buscar variables de nivell 2 que puguin explicar part d'aquesta variància entre grups.
- Quan hi ha un efecte aleatori important en un pendent, hi ha una variació entre grups no explicada de l'efecte de la variable en qüestió. En aquest cas, es poden buscar variables de nivell 2 que ho expliquin, cosa que comportarà que el model tingui "cross-level interactions".
- Cal controlar la necessitat o no d'incloure un cert patró de covariances en els efectes aleatoris.
- En principi, s'ha de ser contrari a deixar efectes no significatius al model, cosa que es podria traduir com a no sobreajustar.

- Cal incloure al model tots els efectes que siguin significatius, a no ser que la base de dades sigui tant gran que alguns efectes, encara que siguin significatius, no siguin importants.
  
- S'ha de tenir en compte els casos següents:
  - a. Cada test que fem per un paràmetre controla per tota la resta d'efectes que estan en el model que fem servir d'hipòtesi nul·la. Com que aquest conjunt d'efectes té una influència tant en la interpretació com en el poder estadístic, els tests resultants dependran d'aquest conjunt.
  
  - b. Estem fent contínuament errors de primer i de segon tipus. Especialment de l'últim, ja que el poder estadístic és sovint baix. Això implica que, el fet trobem que un efecte no és estadísticament significatiu no implica que aquest efecte no existeix en la població. També vol dir que un efecte estadísticament significatiu pot ser resultat de l'atzar, però aquesta probabilitat no és més gran que el nivell de significació, usualment posat a 0.05. Aquest darrer fet és cert quan només realitzem una comparació, ja que en un model amb moltes variables s'està entrant en un terreny de múltiples comparacions amb totes les conseqüències que això comporta: la més important, que la probabilitat d'acceptar almenys una de les comparacions és més gran que 0.05.
  
  - c. Usualment, les anàlisis multinivell estan basades en un nombre limitat de grups. Com que el poder per detectar efectes de les variables de nivell de grup depèn fortament del nombre de grups de les dades, és important tenir molt en compte si tenim un baix poder per les variables a nivell de grup.



Un cop s'han fet tots les anàlisis prèvies i es passa a la construcció del model multinivell, tant Hox (1995) com Snijders & Bosker (1999) recomanen estratègies a seguir. Cal dir que aquesta estratègia està basada en el model lineal, i jo en aquest projecte n'he fet una adaptació al model logístic. Hox (1995) recomana una estratègia que comença amb un model buit i va afegint termes (forward), mentre que Snijders & Bosker (1999) també expliquen a més a més una estratègia que va a la inversa (backward).

L'estratègia *forward* que recomanen comença construint un model pel nivell 1, és a dir, primer s'explica la variabilitat intra-grups, i un cop s'ha fet això es passa a explicar la variabilitat entre grups. Els passos concrets són els següents:

1. Crear un model sense variables explicatives, només amb la constant. Aquest model es coneix com a “Intercept only”, i té a notació següent:

$$y_{ij} = \gamma_{00} + u_{0i} + e_{ij}$$

Un cop s'ha estimat aquest model, tenim una estimació de  $\sigma_u^2$  i  $\sigma_e^2$ , i en un model lineal podríem calcular el percentatge de variabilitat total que és deguda als grups,  $\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$ , per, en les següents etapes, poder avaluar com les variables que introduïm al model (tant de nivell 1 com de nivell 2) fan reduir aquest percentatge. En un model no lineal com el de regressió logística que faré servir, la variància a nivell 1 i la variància a nivell 2 estan en escales diferents i aquesta proporció no es pot calcular. El que sí que es pot fer és comparar directament la variància a nivell de grup en aquest model amb els models successius.

2. Crear un model amb variables del nivell més baix amb coeficients fixos. Valorar la contribució de cada variable explicativa individualment. És a dir, les variables s'entren una a una i es mira si tenen o no significació estadística. En el cas del model lineal es pot mirar amb la diferència de *deviances*, però en regressió logística no, i per tant faré servir el test de

Wald. Posteriorment, s'inclouen totes les variables que han resultat significatives al model, s'avaluen confusions i es deixa el model només amb les variables significatives. En el meu cas he retingut les variables amb un p-valor < 0.10. En aquesta fase, on no s'inclouen variables a nivell de grup, també he treballat paral·lelament amb un model de regressió logística estàndard per comparar si s'arribava al mateix model.

Al final del procés, es pot comparar la variància entre grups obtinguda amb la que teníem al pas 1 per valorar com ha canviat.

3. Valorar si algun pendent té una component de variància significativa entre grups, és a dir, si és aleatori. Aquesta operació pot fer-se coeficient per coeficient, i fins i tot es poden recuperar variables descartades al pas 2. Després de decidir quins coeficients tenen una variància significativa entre grups, es poden afegir totes les components de la variància en un model final i fer el test de si el model final és millor que el model al que s'ha arribat al pas 2.
4. Afegir variables de nivell més alt (nivell de grup), fent que expliquin la constant, és a dir, sense crear interaccions entre variables de diferents nivells. Comprovar si aquestes variables expliquen la variància entre grups, comparant la variància deguda al grup obtinguda amb la que teníem al pas 2.
5. Afegir *cross-level interactions*, entre les variables explicatives de grup i les variables explicatives individuals que tenen el coeficient aleatori, (si no tenen el coeficient aleatori no s'han d'incloure aquestes interaccions, ja que aquestes interaccions intenten explicar la variabilitat en coeficient i no té sentit fer-ho si aquest coeficient és fix).

Partint d'aquest esquema bàsic, es provaran diferents models, traient i afegint variables i avaluant col·linealitats. Les variables amb un p-valor menor de 0.10 es mantindran al model.

### 5.1.2.2. Alternatives

El procés de construir el model primer modelitzant el nivell 1 i seguidament estendre a variables de nivell 2 és elegant però no sempre és el més eficient. Si hi ha variables de nivell 2 o interaccions de variables de diferent nivell que se sap que són importants, es poden incloure al model des de l'inici. Per exemple, es pot esperar que un cert nombre de variables tinguin un pendent diferent a nivell individual que com a variables agregades de grup. En un cas com aquest, pot ser desitjable incloure-les totes dues al model des de l'inici.

Una estratègia alternativa a l'escollida està suggerida a Snijders & Bosker (1999) i consisteix en no distingir entre variables de nivell 1 i nivell 2. Els passos a seguir serien els següents:

1. Seleccionar les variables rellevants de qualsevol nivell en base al coneixement previ. Incloure també interaccions entre variables del mateix i de diferent nivell.
2. Seleccionar d'entre les variables de nivell 1 aquelles en les qual té sentit que hi pugui haver un pendent aleatori.
3. Estimar el model amb els efectes fixos del pas 1 i els aleatoris del pas 2.
4. Fer el test de significació dels pendents aleatoris i els que no ho siguin tractar-los com a fixos.
5. Fer el test de significació dels pendents, i excloure els que no són significatius. Aquest pot ser un bon moment per considerar la inclusió d'altres interaccions.

6. Comprovar que les variables que inicialment no s'havien inclòs per considerar-les poc plausibles, realment serien no significatives. En cas de ser-ho, s'inclouen al model.

Un cas extrem del pas 1 és incloure totes les variables disponibles al model, més un gran nombre d'interaccions, i a més a més, posar pendents aleatoris a totes les variables de nivell 1. Això significaria una estratègia *backward*, on es comença amb un model molt gran i es redueix pas a pas traient les variables no significatives. A cada pas es podria valorar si variables excloses poden tornar a entrar al model, cosa que significaria una estratègia *stepwise*. Quan tenim moltes variables aquesta estratègia és pràcticament impossible d'aplicar. De tota manera, ha de quedar clar que es podria arribar a un model final de moltes maneres, i no necessàriament s'arribaria al mateix. A més, és ben possible que diferents models s'ajustin bé a unes dades, i que no hi hagi raons concloents per preferir-ne l'un o l'altre. En aquests casos, és millor acceptar aquesta indeterminació i deixar-ho per resoldre en noves investigacions que no fer una tria sense arguments de pes entre els models per quedar-se'n un de sol com a bo.

Els tractaments que s'han exposat per l'especificació del model són molt guiats per les dades. Si es té una hipòtesi a priori, és útil distingir entre els paràmetres en els quals es basa la hipòtesi i les altres parts del model, que són necessàries per obtenir un bon ajust i en conseqüència un test vàlid per la hipòtesi. Per aquesta última part es pot fer el tractament guiat per les dades, mentre que els paràmetres pels quals es fa el test han de romandre en el model sempre.

## 5.2. Estratègia de validació

En un model de més d'un nivell tenim residus a cada un dels nivells, i per tant aquesta anàlisi no es pot limitar a l'anàlisi dels residus individuals. L'anàlisi de residus en models multinivell està descrit en algun llibre, com el de Snijders & Bosker (1999) o el de Goldstein (1995), però els dos es basen en el model lineal i no en el logístic. El

manual del paquet MlwiN dóna alguns exemples de possibles gràfics de residus en un model logístic multinivell, però no explica amb detall quins passos s'ha de seguir, quins són els bons gràfics a mirar, què s'hi ha de buscar, etc.

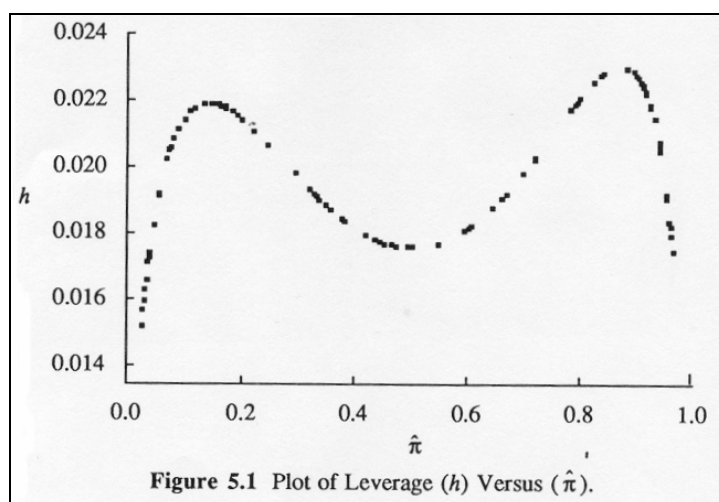
### **5.2.1. Residus a nivell individual**

Els residus del nivell 1 que obtenim són comparables als residus d'un model logístic estàndard. Així doncs es poden aplicar les mateixes tècniques de validació de residus, encara que degut a que el mètode d'estimació no és el mateix, alguns diagnòstics no es poden calcular, com per exemple el canvi en la Deviance degut a eliminar una observació suggerit a Hosmer i Lemeshow (1989). Hosmer i Lemeshow (1989) és el llibre que tracta més extensament el tema de l'anàlisi de residus en models logístics, i té un capítol dedicat estrictament a això. Així doncs, per l'anàlisi dels residus a nivell 1 em basaré en aquest llibre.

Les mesures de diagnòstic d'un model estan destinades a valorar si l'ajust del model és "suportat" per tots els patrons covariable de les dades. En un model lineal una assumptió clau és que la variància residual no depengui del valor previst, però en regressió logística tenim errors binomials, i per tant, la variància residual és funció del valor previst. A més, en un model logístic es poden calcular diversos tipus de residus, com per exemple els residus deviance, o els residus de Pearson. Els residus es poden estandarditzar per treballar amb els residus estandarditzats que poden ser més útils. En un model de regressió logística sense tenir les dades agregades, el gràfic de residus contra valors previstos que és tant útil en la majoria de models no ens aporta informació i per tant no es farà servir.

A part dels residus es poden calcular altres quantitats per cada patró covariable, com per exemple el leverage. En un model lineal el leverage és proporcional a la distància de les  $x$ 's del patró covariable en qüestió a la mitjana de les  $x$ 's de totes les dades, i és important perquè ens indica dades que són diferents a la resta i que a més poden tenir una influència considerable en els valors estimats dels paràmetres. En

regressió logística el comportament del leverage és diferent. A la següent figura es pot apreciar l'exemple que donen Hosmer i Lemeshow. En ella es pot observar que el leverage creix a mesura que la probabilitat estimada s'allunya de 0.5, que és la mitjana de les  $x$ 's en el seu exemple, però quan les probabilitats esdevenen menors de 0.1 o majors de 0.9 el leverage decreix ràpidament i tendeix a zero. Aquest exemple mostra que els punts més extrems en la distribució de les  $x$ 's són els que tenen el leverage més petit, just el contrari del que passa en el cas lineal. La conseqüència d'aquest fet és que, per interpretar correctament el leverage en models logístics, cal saber si la probabilitat és molt petita ( $<0.10$ ) o molt gran ( $>0.90$ ). Si la probabilitat està a l'interval  $(0.10, 0.90)$  el leverage es pot pensar com una distància. En cas contrari, el valor del leverage pot no mesurar una distància en el sentit que els valors més allunyats de la mitjana tinguin un leverage més alt.



Una altra mesura de diagnòstic útil és mesurar l'efecte que té esborrar les observacions amb un patró covariable determinat en el valor de les estimacions dels paràmetres, mesura que anomenarem influència. Aquest canvi és l'anàleg a la mesura d'influència anomenada DFITS per models multinivell (Goldstein 1995). Degut a que aquesta quantitat depèn del leverage, es pot deduir que tindrà valors alts quan la probabilitat estimada estigui entre 0.1 i 0.3 (o entre 0.7 i 0.9), mentre que tindrà valors petits entre 0 i 0.1 (o entre 0.9 i 1) i valors moderats a l'interval que va entre 0.3 i 0.7. Aquesta mesura és molt útil ja que ens permet identificar els patrons covariable que

tenen una alta influència en els valors dels paràmetres estimats, i un com identificats, poder començar a investigar el rol que tenen en l'anàlisi.

El valors grans o petits que ha de prendre el leverage o la influència segons les probabilitats previstes que donen Hosmer i Lemeshow són els esperats, i no els que poden sortir en un exemple concret. Per tant, només s'han d'usar com a guia per entendre i interpretar les mesures de diagnòstic.

En regressió logística s'ha de confiar bàsicament en una valoració visual dels diagnòstics, ja que la distribució dels diagnòstics un cop ajustat el model és coneguda només en un nombre limitat de casos. Hosmer i Lemeshow recomanen alguns gràfics que són els més fàcilment calculables i els que tenen una interpretació més senzilla. Els que tenen quantitats que es poden obtenir amb els nostres models són el gràfic de la influència contra la probabilitat prevista i el de la influència contra el leverage. El primer és el que té més importància, mentre que el segon ajuda a veure com contribueix el leverage a la influència.

En el gràfic d'influència contra la probabilitat prevista hem de buscar punts que tinguin valors d'influència alts i separats de la resta de punts. Identificarem els punts en els quals passi això i els estudiarem, mirant quines són les característiques que els poden fer especials i estudiant què passa si els traiem del model. Cal tenir present que la influència estudia com canvien tot el conjunt de coeficients, i no el coeficients individual d'una variable. El segon gràfic ens pot ajudar a valorar si la influència és deguda o no al leverage.

## **5.2.2. Residus a nivell de centre**

Pel que fa als residus a nivell 2, hem suposat (capítol 4) que segueixen una distribució normal centrada a zero i amb una certa variància. Aquests residus són comparables als residus de nivell 2 d'un model lineal jeràrquic, i per tant en aquest cas sí que podrem seguir l'anàlisi recomanat en els llibres de models multinivell.

Cal tenir en compte que els residus estimats a nivell 2 estan influenciats pels residus estimats a nivell 1. És per això que sempre s'han de mirar sempre primer els residus a nivell 1.

Els residus a nivell dos són els efectes aleatoris  $u_{0i}$ . Aquests són variables latents i no paràmetres estadístics, i per tant no s'estimen quan es fa l'estimació dels paràmetres. No obstant, es poden estimar pel mètode conegut com a *empirical Bayes*, que produeix el que es coneixen com a *posterior means*. La idea bàsica d'aquest mètode és que  $u_{0i}$  s'estima combinant dos tipus d'informació:

- a) les dades del grup  $i$ -èssim
- b) l'assumpció que les  $u_{0i}$  no observades són una variable aleatòria amb distribució normal amb mitjana 0 i variància  $\sigma_{u_0}^2$ .

Dit d'una altra manera, la informació de les dades es combina amb la informació de la població. La idea està treta de l'estadística bayesiana. Bàsicament es combina la informació prèvia (prior) que està basada en la població de la que extraiem la mostra, i el coneixement posterior, que està basat en les observacions del grup.

Així doncs, una de les assumpcions que fem respecte als residus de nivell 2 és que segueixen una distribució normal, i per tant haurem de verificar aquesta hipòtesi. Això ho podem fer amb un histograma o amb un gràfic de normalitat.

També, per les variables a nivell de centre, es faran gràfics per avaluar la relació  $\beta_{0i} = \gamma_{00} + \gamma_{01} z_i + u_{0i}$ , descrita al capítol 4, que relaciona les constants de cada centre amb una variable a nivell de centre  $z_i$ . Si ens fixem en aquesta equació veiem que es correspon a un model de regressió lineal simple entre les constants de cada centre,  $\beta_{0i}$ , i la variable a nivell de centre  $z_i$ , i els residus d'aquesta relació són  $u_{0i}$ . Així doncs, per avaluar gràficament aquesta relació, podem fer un gràfic de les constants del model abans d'incloure-hi la variable  $z_i$  contra aquesta variable. Aquest gràfic pot ajudar a veure si existeix relació i a valorar si aquesta és lineal o no, si hi ha un punt a partir del



qual hi ha un graó i per tant la variable s'hauria de categoritzar, si la relació és només deguda a un centre específic, etc.

A nivell 2 el gràfic de residus estandarditzats contra valors previstos sí que és útil. Els residus són les pertorbacions a nivell de centre, i com a valors previstos podem prendre les mitjanes per centre de la part fixa del model,  $\gamma_{00} + \sum_{p=1}^p (\gamma_{p0})' X_{pij} + \sum_{q=1}^q \gamma_{0q}' Z_{qi}$ . Com és habitual en els gràfics de residus contra valors previstos, hi haurem de buscar que el conjunt de punts no tingui forma, que l'esperança sigui zero per tots els valors previstos i que la variabilitat sigui constant.

També es poden calcular diagnòstics a nivell 2. Es calcularà la influència i el leverage de cada un dels centres i s'avaluarà gràficament quins centres són els que tenen més influència en l'anàlisi. Un cop identificats s'estimarà el model sense incloure'ls i es compararan els resultats amb els anteriors.

## 5.3. Estratègia d'interpretació

### 5.3.1. Interpretació dels resultats

Al capítol 6 es fa l'anàlisi exploratòria de dades. En aquesta anàlisi es descriuen les variables i també les relacions que existeixen entre algunes variables dos a dos. Aquestes relacions són les que s'observen realment a la nostra mostra si creuem les variables en qüestió, sense corregir l'efecte per cap altra variable. En canvi, al capítol 7 es duu a terme la modelització, i totes les relacions que obtenim estan ajustades per la resta de variables incloses al model. Per les variables categòriques, l'odds ratio per una categoria particular és el canvi en els odds que es produeix al passar de la categoria de referència a la categoria en qüestió mantenint fixes la resta de covariables del model. Per les variables contínues, l'odds ratio és l'increment de l'odds de patir la malaltia que es produeix a l'augmentar en una unitat la variable mantenint constants la resta. Les

variables a nivell de centre s'interpreten com qualsevol variable a nivell individual, és a dir, el seu efecte és en la variable resposta individual, i no en el percentatge del centre. Per tant, per una variable expressada en percentatge, l'odds ratio és l'increment en els odds de la malaltia que es produeix a l'augmentar un 1% el percentatge del centre de la variable que s'està estudiant, mantenint la resta de variables del model constants. Per interpretar els efectes de les variables ho farà una per una, però sempre caldrà tenir en compte que l'efecte calculat és el que s'obté si la resta de variables romanen constants.

### **5.3.2. Interpretació dels odds ratio**

Quan tenim una variable resposta dicotòmica es poden calcular diversos estadístics per mesurar la relació que té amb una variable explicativa. Suposem que la variable explicativa és també dicotòmica i indica l'exposició o no a un determinat factor de risc, i que la variable resposta indica la presència o no d'una malaltia. Una de les mesures d'associació més comunes i més fàcils d'entendre és el risc relatiu (RR), que és el quocient de la probabilitat de tenir la malaltia dels exposats i dels no exposats. Si denotem per  $P(M)$  la probabilitat de patir la malaltia, per  $E^+$  els exposats i per  $E^-$  els no exposats, el risc relatiu es calcula com a  $RR = P(D|E^+)/P(D|E^-)$ .

Cal dir que el risc és la probabilitat de desenvolupar una malaltia durant un cert període de temps. Per tant, la denominació "risc" només és aplicable quan tenim un estudi en el qual partim d'una població sana, la seguim durant un temps i al final avaluem quantes persones han adquirit la malaltia. En el cas d'un estudi transversal com el nostre,  $P(D|E^+)/P(D|E^-)$  no estima un risc de malaltia sinó una probabilitat relativa d'escollir aleatòriament una persona que tingui la malaltia en el moment de l'estudi. És per això que, tot i que la fórmula és la mateixa, en els estudis transversals no és correcte anomenar el resultat "risc relatiu" sinó que s'ha d'anomenar "prevalença relativa" o quocient de prevalences.

L'odds ratio (OR) és una altra de les mesures d'associació possibles. L'odds de patir la malaltia és el quocient de la probabilitat de tenir la malaltia i la probabilitat de no tenir-la, és a dir,  $\text{odds} = P(D)/(1-P(D))$ . L'odds ratio és el quocient de l'odds de patir la malaltia en els exposats respecte a l'odds de patir la malaltia en els no exposats. És a dir:

$$\text{OR} = \frac{P(D|E+) / (1 - P(D|E+))}{P(D|E-) / (1 - P(D|E-))}$$

L'odds ratio és la mesura que es fa servir en els estudis cas-control perquè en aquests estudis el risc relatiu no és calculable, ja que la proporció de casos respecte a controls està fixada per l'investigador. El que ha passat a la pràctica és que l'odds ratio s'ha acabat usant en altres dissenys, com els estudis transversals. Això bàsicament és degut a la facilitat d'ajust dels models de regressió logística, que permeten obtenir les relacions ajustades per altres variables i que tenen la propietat que, si exponenciem els coeficients obtinguts amb aquest model, obtenim odds ratios. Com que la majoria de paquets estadístics porten implementada aquesta tècnica i tradicionalment s'ha explicat que no hi ha un model que permeti ajustar les relacions per altres variables i reporti riscos relatius, l'ús dels odds ratios ha esdevingut la pràctica habitual en estudis transversals, sobretot en l'àmbit de la medicina.

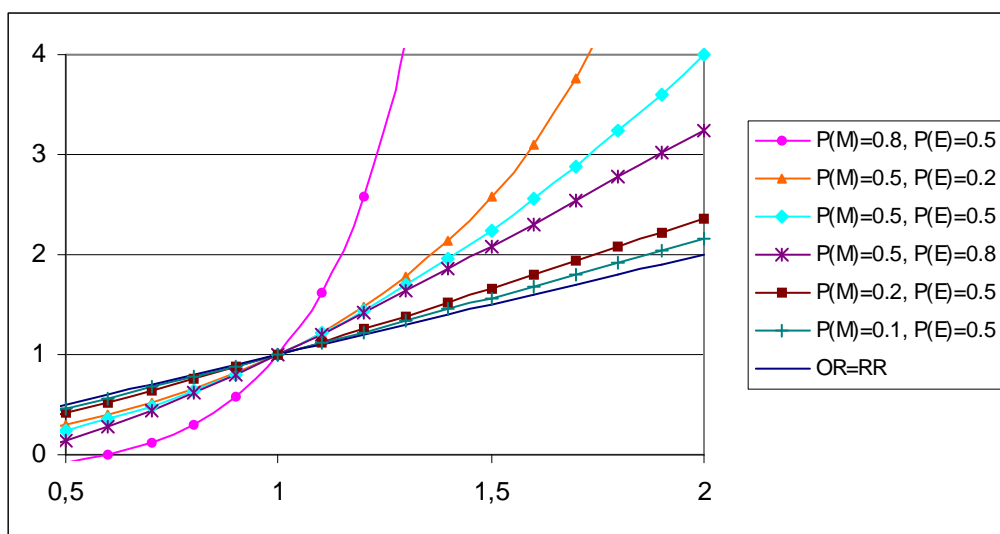
No obstant, mitjançant models lineals generalitzats, especificant la família binomial i el link log, l'exponenciació dels coeficients dona lloc a riscos relatius o prevalences relatives. Per més informació sobre aquests models, es pot consultar Wacholder (1986) o Zocchetti (1995). No tots els paquets estadístics permeten especificar el link log amb la família binomial, però per exemple Stata o SAS ja permeten fer-ho.

Els odds ratio són una magnitud difícil d'interpretar, ja estem acostumats a pensar en probabilitats i no en odds. El que fan la majoria d'estudis és fer servir una propietat per la qual, quan la prevalença de la resposta és petita (normalment es diu menor d'un 10%) l'odds ratio i el risc relatiu són molt semblants. Així, quan la prevalença és petita, l'odds ratio es pot acabar interpretant com si fos un risc relatiu.

Zocchetti (1997) ensenya la relació que hi ha en una taula 2x2 entre l'odds ratio i el quocient de prevalències per diferents prevalències de la malaltia i de l'exposició. La relació entre ambdues quantitats és la següent:

$$OR = RR * \frac{1 - P(E) + RR * P(E) - P(M)}{1 - P(E) + RR * P(E) - RR * P(M)}$$

i fent un gràfic de la relació per diferents prevalències de la malaltia i de l'exposició obtenim la figura que es mostra a continuació, on es pot veure que el canvi en la prevalença de la malaltia afecta més l'OR que el canvi en la prevalença de l'exposició, i que, quan la malaltia és rara ( $P(M) < 0.10$ ), no hi ha moltes discrepàncies entre l'OR i el RR. Aquest últim fet es coneix com a la hipòtesi de la malaltia rara, i ens habilita a poder interpretar els odds ratios com a riscos relatius o quocients de prevalències. Cal observar, però, que quan la prevalença de la malaltia i de l'exposició són del 50%  $OR = RR^2$ , i cometem un error molt gran si interpretem un odds ratio de 4 com un risc relatiu de 4, quan en realitat és de 2.



En el cas d'aquest projecte, la prevalença d'asma és petita (8.4%), i la hipòtesi de la malaltia rara es compleix. Per tant, els odds ratio aproximaran els quocients de prevalències i els interpretarem com si ho fossin.

# Capítol 6. Anàlisi exploratòria de les dades

Seguint l'estratègia explicada al capítol 5, l'anàlisi exploratòria de dades es farà en diverses fases. Primer estarà centrat en les variables a nivell individual i després es miraran les variables a nivell de grup. A més, en els dos grups es separarà les variables específiques que es volen estudiar en el projecte, que són les socioeconòmiques, i la resta de variables que serveixen per ajustar la relació.

## 6.1 Descripció de les variables individuals per centre

### 6.1.1. Participació i variable asma

A la **taula 1** es poden veure els 32 centres participants, amb els 15 països als quals pertanyen. La primera columna mostra els efectius que tenim a la mostra final. Tal i com s'ha explicat a la secció 2.3, el mostreig de l'ECRHS estava compost per dues fases. En la primera, mitjançant un marc de mostreig s'enviava un petit qüestionari de poques preguntes referents a la salut respiratòria a una mostra aleatòria d'individus. Als que no responien se'ls tornava a enviar varies vegades el qüestionari i finalment se'ls trucava per intentar aconseguir una resposta. Les grandàries de mostra per cada centre variaven, però la mostra inicial havia de contenir com a mínim 1.500 homes i 1.500

doncs. La resposta a aquesta fase va ser diferent entre centres. Posteriorment, la fase 2 s'escollia una mostra aleatòria més petita, un 20% dels participants a la fase 1, se'ls passava un qüestionari molt més llarg. La resposta a aquesta fase també va ser diferent per centres, i el nombre total d'individus que van respondre és el que mostra la primera columna de la **taula 1**. La no-resposta és un factor que s'ha estudiat molt en l'estudi ECHRS, i en els estudis en els quals s'ha tingut en compte, les conclusions pràcticament no han variat (ECHRS 1996, Janson 1997). A la mostra escollida se'ls feien, a part del qüestionari, proves funcionals i extracció de sang si hi donaven consentiment. En aquest projecte necessitem les dades d'IgE en sang per valorar la sensibilització a diferents al·lèrgens, i per aquesta raó perdem molts participants que no van donar el seu consentiment a participar en aquesta part. La segona columna de la **taula 1** mostra els participants de cada centre que sí que tenen mostra de sang. L'últim factor que falta que ens fa perdre observacions són els valors mancants en altres variables que es fan servir en el projecte. La tercera columna de la **taula 1** mostra el nombre d'individus finalment utilitzats. Tenim dades completes per 10.971 (58%) individus. Els valors mancants són deguts bàsicament a la gent que no donava consentiment per fer l'extracció de sang per mesurar la IgE. El percentatge de dades completes si no es consideren les variables de IgE és del 91%. Si ens fixem en els individus perduts a causa de no tenir mostra de sang o tenir valors mancants a altres variables, tenim quatre centres amb més del 50% de perduts, que són Barcelona, Portland, Antwerp city i Hawkes Bay. D'altra banda, hi ha dos centres amb menys del 10% de pèrdues: Albacete i Erfurt.

L'última columna de la **taula 1** mostra la prevalença d'asma en l'actualitat pels diferents centres. La variació entre centres és considerable (rang: 2.80%-15.72%), i és estadísticament significativa amb el test de Chi-quadrat ( $p < 0.001$ ). Cal observar que dins d'un mateix país la diferència també pot ser gran. Per exemple, a Espanya el rang va del 4.00% al 11.06%, i si fem el test de Chi-quadrat només en els centres espanyols, les diferències també assoleixen significació estadística ( $p = 0.005$ ). El mostreig és aleatori i per tant un biaix de selecció no pot ser la causa de la variació en la prevalença. La no-resposta sí que podria tenir-hi un efecte, però aquest factor ha estat bastant estudiat. Com que es tenien dades del marc de mostreig respecte a sexe i edat, a ECHRS (1996) es van desenvolupar models per predir els símptomes dels que no responien

utilitzant aquestes variables més el nombre de cartes que s'havien enviat a cada individu fins que contestava, l'estació de l'any en la qual se li havia enviat i el centre. Després es van estimar les prevalences en la fase 1 tenint en compte l'efecte de la no-resposta i aquest només suposava una petita font de variació comparat amb la variació total que hi ha en l'estudi. També es tenien dades dels que van respondre a la fase I però no van respondre a la fase 2, i es va repetir el procediment amb idèntic resultat. En tots els articles on s'ha ajustat per no-resposta, el rànquing de països no s'ha vist afectat i l'efecte que s'estudiava pràcticament no ha variat. Una altra hipòtesi que es va plantejar va ser que les variacions fossin degudes als diferents idiomes en els que estava el qüestionari. Es va veure que els països de parla anglesa eren els que tenien una prevalença més alta i per això es va formular aquesta hipòtesi. Verlato (2002) però, va notar que això podia ser degut a que tots aquests centres tenen clima oceànic, que té característiques com l'alta humitat, que afavoreixen els símptomes. A més, s'observen grans variacions en centres que fan servir el mateix idioma.

D'altra banda, és perfectament plausible i no sorprèn que hi hagi variació, degut a que l'asma té una component ambiental molt forta i en els centres participants hi ha un gran espectre de climes, nivell de contaminació, i nivell d'al·lergens.

Centre	País	N	N amb mostra de sang	N sense missing†	Prevalença (%) d'asma en l'actualitat
South Antwerp	Bèlgica	558	327	304	5.26
Antwerp city	Bèlgica	564	267	244	9.02
Hamburg	Alemanya	1252	775	769	7.02
Erfurt	Alemanya	731	718	700	3.43
Barcelona	Espanya	393	175	162	5.56
Galdakao	Espanya	486	365	325	4.00
Albacete	Espanya	435	392	392	10.98
Oviedo	Espanya	357	244	246	8.50
Huelva	Espanya	271	215	200	11.06
Grenoble	França	473	430	395	5.82
Dublin	Irlanda	454	248	239	12.55
Pavia	Itàlia	310	250	225	8.89
Turin	Itàlia	244	177	159	15.72
Verona	Itàlia	340	332	298	6.38
Groningen	Holanda	380	354	295	5.08
Bergen op Zoom	Holanda	452	418	406	7.14
Geleen	Holanda	415	388	336	5.95
Cambridge	Regne Unit	277	187	173	9.94
Cardiff	Regne Unit	381	321	309	13.92
Ipswich	Regne Unit	448	365	345	13.04
Norwich	Regne Unit	473	363	337	12.76
Reykjavik	Islàndia	562	516	428	2.80
Bergen	Noruega	835	749	664	6.17
Goteborg	Suècia	682	574	506	8.10
Umea	Suècia	552	425	375	9.07
Uppsala	Suècia	622	524	476	7.98
Wellington	Nova Zelanda	481	314	280	13.57
Christchurch	Nova Zelanda	457	270	239	13.81
Hawkes Bay	Nova Zelanda	316	142	140	12.14
Portland	Estats Units	723	343	311	11.58
Melbourne	Austràlia	669	520	494	12.96
Tartu	Estònia	431	349	294	6.46
Total		18.790	12.033	10.971	8.36
Test de $\chi^2$					166 (31 g.ll.) p<0.001

† Sense missing a les variables: asma; variables d'IgE ; edat; sexe; història familiar d'asma; classe social; educació; exposició laboral a pols biològica, mineral o gas i fums; infeccions respiratòries abans dels 5 anys d'edat; nombre de germans; tabaquisme; presència de verdet a la casa i tabaquisme passiu.

**Taula 1. Centres participants, mostra total, efectius sense valors mancants, i prevalença d'asma en l'actualitat per cada centre.**



## 6.1.2 Variables socioeconòmiques

La **taula 2** descriu les variables socioeconòmiques a nivell individual que es fan servir en l'estudi, estratificades per centre. Així, podem observar-hi els percentatges d'individus en cada una de les classes socials, en cada nivell d'educació, i el percentatge d'exposats a pols biològica o mineral, gasos o fums. El percentatge total d'individus a la classe social alta és del 36.8%. Aquest percentatge varia des del 18% de Galdakao fins al 60% de Melbourne, i podem trobar diferències grans entre centres del mateix país. Les classes intermitges (III manual i no manual) no tenen tanta variació i van del 9% al 33%, i curiosament el màxim i el mínim estan a Itàlia pels no manuals. Les classes socials baixes presenten més variació, que va del 4% a Tartu fins al 39% de Verona, i la mitjana és del 15.5%. Els no classificables són un 5.9% del total, que van des dels zero de Goteborg, Melbourne i Antwerp city fins al 17% d'Albacete.

La variable educació mostra més variabilitat entre centres, cosa que podríem considerar lògica ja que en els diferents països hi ha diversos tipus de sistemes educatius. El percentatge d'individus que deixen els estudis abans dels 16 anys és bastant petit (13.6%). En aquestes dades destaca Espanya amb un percentatge molt elevat en tots els seus centres, amb una mitjana del 36%. Cal tenir en compte que els participants van néixer de l'any 48 al 72, on a Espanya no era molt freqüent seguir amb els estudis i l'escolarització obligatòria arribava només fins als 14 anys. Itàlia i Irlanda són els altres dos països que tenen algun centre amb un percentatge superior al 25%. Pel que fa als que finalitzen els seus estudis més tard dels 20 anys, els percentatges oscil·len del 12% de Hawkes Bay al 66% de Portland, amb una mitjana del 41.8%. En total, són més d'un 86% els que estudien fins més enllà dels 15 anys.

Els qui treballen en una feina en la qual tenen exposició a pols mineral o biològica, gasos, o fums, són un 13.2%. El rang no és molt ample i va del 5% de Cambridge al 23% de Umea. Espanya presenta també dos centres (Galdakao i Albacete) amb un percentatge més gran del 20%.

Aquestes tres variables socioeconòmiques donen una idea de les diferències que existeixen en les diferents ciutats de l'estudi, i de fet, els percentatges per centre d'aquestes variables es faran com a variables a nivell de centre, fent d'indicadors socioeconòmics de les ciutats.

Centre	País	Classe social					Educació (edat al finalitzar)			Alta exposició laboral a contaminants
		I-II	III no manual	III manual	IV-V	No class.	<16	16-19	≥20	
South Antwerp	Bèlgica	39	31	13	15	2	9	40	50	9
Antwerp city	Bèlgica	36	27	15	22	0	11	34	55	13
Hamburg	Alemanya	45	13	17	13	12	8	36	56	9
Erfurt	Alemanya	41	13	31	9	6	0	62	38	10
Barcelona	Espanya	36	23	22	17	2	32	33	35	13
Galdakao	Espanya	18	20	27	24	11	37	37	26	21
Albacete	Espanya	21	14	25	22	17	44	38	19	22
Oviedo	Espanya	38	21	17	18	6	26	33	41	13
Huelva	Espanya	20	26	18	23	13	39	34	27	12
Grenoble	França	51	28	10	10	2	7	33	60	8
Dublin	Irlanda	25	21	26	24	5	29	54	16	16
Pavia	Itàlia	28	29	18	12	12	27	39	34	9
Turin	Itàlia	31	33	18	11	7	26	36	37	11
Verona	Itàlia	24	9	20	39	8	28	45	27	10
Groningen	Holanda	50	23	16	9	2	10	35	55	10
Bergen op Zoom	Holanda	31	24	27	16	2	17	58	25	19
Geleen	Holanda	35	20	23	14	8	17	46	38	14
Cambridge	Regne Unit	59	19	11	9	1	8	45	47	5
Cardiff	Regne Unit	29	30	21	19	1	23	61	16	13
Ipswich	Regne Unit	29	24	19	15	13	21	64	15	13
Norwich	Regne Unit	31	20	22	17	11	19	58	23	16
Reykjavik	Islàndia	31	27	24	13	5	8	32	60	14
Bergen	Noruega	44	22	14	14	5	2	33	66	11
Goteborg	Suècia	22	26	30	22	0	4	48	48	15
Umea	Suècia	31	19	24	24	2	7	46	47	23
Uppsala	Suècia	32	12	24	18	14	3	32	64	19
Wellington	Nova Zelanda	45	23	17	13	3	14	56	30	9
Christchurch	Nova Zelanda	36	23	26	11	5	14	66	21	15
Hawkes Bay	Nova Zelanda	37	28	14	20	1	17	71	12	19
Portland	Estats Units	52	20	17	9	1	2	33	66	9
Melbourne	Austràlia	60	22	11	7	0	8	48	45	8
Tartu	Estònia	44	16	33	4	2	2	53	45	17
Total (rang)		36.8 (18-60)	21.0 (9-33)	20.8 (10-33)	15.5 (4-39)	5.9 (0-17)	13.6 (0-44)	44.7 (32-71)	41.8 (12-66)	13.2 (5-23)

Taula 2. Percentatges de les variables socioeconòmiques per centre.

Podem observar a la **taula 3** com estan relacionades les tres variables que mesuren el nivell socioeconòmic. L'alta exposició laboral a contaminants té una relació clara amb la classe social: si bé en les classes socials més altes aquesta exposició és mínima ja que no arriba ni al 2%, en les classes III manual i IV-V més d'un 30% hi estan exposades. L'exposició laboral a contaminants també creix a mesura que disminueix el nivell d'educació: és d'un 6% en els d'educació alta, d'un 16% en els d'educació mitjana i d'un 25% en els d'educació baixa. D'altra banda també existeix una relació clara entre nivell d'estudis i classe social. En els de classe social més alta, un 65% han assolit un nivell d'estudis alt, mentre que aquest percentatge en els de classe IV-V és només del 20%. De la mateixa manera, el percentatge d'individus de baixa educació en els de classe social I-II és un 5% i va creixent gradualment a mesura que disminuïm la classe social fins a arribar a un 27% a les classes IV-V. Els que no es van poder classificar per classe social no sembla que tinguin un patró de nivell d'estudis similar a cap de les categories de classe social. Així doncs, aquestes dades van en contra de poder afirmar que els no classificables són realment de classe baixa.

		Alta exposició Laboral a contaminants					
Classe social							
	I-II	1.2%					
	III no-manual	1.9%					
	III manual	35.2%					
	IV-V	32.6%					
	No classificable	0%					
Educació		Classe social					
		I-II	III no-manual	III manual	IV-V	No classificable	
	Alta	6.4%	64.6%	34.7%	23.8%	19.8%	45.0%
	Mitjana	16.0%	30.6%	54.9%	56.7%	53.1%	31.5%
	Baixa	24.9%	4.8%	10.4%	19.5%	27.1%	23.5%

**Taula 3. Relació entre les variables alta exposició laboral a pols biològica o mineral, gas o fums, classe social i educació.**

### 6.1.3 Variables confusores

A les **taules 4a i 4b** es descriuen les altres variables a nivell individual que es fan servir en el projecte per a ajustar la relació, que són les llistades a la secció 3.4.2. Els percentatges o les mitjanes d'aquestes variables també es mostren estratificades per centre.

El seleccionats tenen una mitjana de 33.7 anys, amb un rang de mitjanes per país que va dels 31 als 37. El percentatge de dones és pràcticament del 50%, encara que si ho mirem per centres, n'hi ha algun que està més desequilibrat. Per exemple, Cambridge té un 63% de dones. De fet en tots els centres del Regne Unit el percentatge de dones és superior al d'homes. Pel cantó contrari, Dublin té un 56% d'homes i diversos centres, entre ells tres centres espanyols, tenen un 55% d'homes. El pare, la mare, o els dos d'un 11.4% dels participants ha tingut asma. Aquest percentatge s'enfila fins al 20% a Austràlia o al 18% a Uppsala, pertanyents a països on ja se sap que hi ha alts índexs d'asma. Els percentatges del Regne Unit, un país que també té una alta prevalença d'asma, són més moderats. Aquesta variable no sembla tenir variacions molt grans dins d'un mateix país.

Les variables de tabaquisme són bastant heterogènies. El percentatge de persones que no han estat mai fumadores va del 29% al 61% depenent del centre, amb una mitjana del 42.2%. Així doncs, més de la meitat de la població ha estat fumadora en algun moment, i un 36% ho és encara actualment. Nova Zelanda, el Regne Unit i Estats Units destaquen en ser els països amb més persones que no han fumat mai. A l'altra banda es posicionen Holanda, Alemanya, Islàndia i Espanya, qui a més a més té els índex d'ex-fumadors més baixos, tret del centre de Barcelona. Els fumadors de més d'un paquet al dia són un 14.1%, i el rang entre centres va del 4% al 31%. En aquesta categoria Espanya és la que té un percentatge més alt amb un 28%, mentre que Nova Zelanda en té només un 6%. La gent que afirma estar exposada al fum del tabac és un 57% en el total de la mostra, i el rang per centres és bastant ampli (35%-84%). En concordància amb el nombre de fumadors actius, també és a Espanya on hi ha més tabaquisme passiu.

La mitjana de germans és 1.3, un valor que oscil·la des dels 0.7 de Hamburg fins als 2.9 d'Irlanda. Un 10% de la població va tenir infeccions respiratòries abans del 5 anys (**taula 4b**), i el rang entre centres va del 2% al 16%.

Pel que fa a la sensibilització a al·lèrgens, mesurada a través de la IgE en sang, la més comuna és la sensibilització als àcars de la pols (Der p1) i a la herba timotea, ambdues presents en un 19% de la població. Els més sensibilitzats a Der p1 es concentren per països, i Nova Zelanda, Irlanda i Austràlia són els que en tenen més amb valors pròxims al 30%, mentre que Islàndia i Suècia són els que en tenen menys, al voltant 9%. Novament els països de parla anglesa (Estats Units, Regne Unit, Nova Zelanda, Austràlia) són els més sensibilitzats a herba timotea, i també a gat. Per bé que això pot fer pensar que hi ha un fort component genètic, Burney (1997) en una anàlisi que incloïa més centres que aquest projecte va concloure que les variacions trobades no eren degudes majoritàriament a l'heterogeneïtat genètica, ja que la distància genètica entre els centres no és suficientment gran com per explicar les grans diferències, i a més a més es trobaven també diferències grans dins d'un mateix país. En aquest article, Burney es decanta més perquè les diferències siguin degudes a diferències en l'ambient i en l'estil de vida. Pel que fa al *cladosporium*, Barcelona amb un 12% de sensibilitzats està molt allunyat de la resta, fins i tot dels altres centres espanyols, com per exemple Albacete que no té cap sensibilitzat a aquest al·lèrgè.

La rinitis afecta a una quarta part de la mostra, i la seva variació és gran (rang 12%-47%). Els països amb una alta prevalença de rinitis coincideixen, com era d'esperar ja que la rinitis és una malaltia al·lèrgica, amb els de més alta sensibilització a al·lèrgens. Per últim, la pregunta sobre si ha tingut verdet a casa en l'últim any la van respondre afirmativament el 22% de la mostra. Els percentatges oscil·len des d'aproximadament un 5% a Islàndia i Noruega fins a més del 50% a Nova Zelanda i Austràlia.

Amb aquestes taules queda patent, doncs, que les variables confusores individuals també presenten una variabilitat considerable entre els centres que componen la mostra.

Centre	País	Edat (mitjana)	Sexe: dona	Història familiar d'asma	Tabaquisme				Tabaquisme passiu	Nombre de germans (mitjana)
					No	Ex	<20 cigs.	≥20 cigs.		
South Antwerp	Bèlgica	33.97	47	9	45	26	15	14	65	1.54
Antwerp city	Bèlgica	33.35	52	8	37	22	22	19	70	1.27
Hamburg	Alemanya	32.91	45	9	32	24	20	23	60	0.73
Erfurt	Alemanya	32.47	48	8	36	21	27	16	49	0.94
Barcelona	Espanya	33.43	45	8	29	21	25	25	75	1.27
Galdakao	Espanya	32.44	45	14	34	13	27	25	84	1.56
Albacete	Espanya	32.33	53	10	32	13	25	31	71	1.62
Oviedo	Espanya	35.17	50	16	29	16	25	31	65	1.33
Huelva	Espanya	32.96	45	12	35	11	29	26	83	1.36
Grenoble	França	36.53	47	8	42	29	19	9	45	1.21
Dublin	Irlanda	32.36	44	9	38	18	24	21	71	2.85
Pavia	Itàlia	34.59	47	12	41	18	20	20	69	0.93
Turin	Itàlia	34.19	48	13	47	23	16	14	67	0.87
Verona	Itàlia	32.86	51	9	44	22	23	10	62	1.23
Groningen	Holanda	35.2	51	10	33	17	29	21	71	1.54
Bergen op Zoom	Holanda	33.86	50	10	36	25	28	11	71	1.56
Geleen	Holanda	34.57	49	8	36	25	28	12	68	1.75
Cambridge	Regne Unit	33.56	63	15	61	19	16	4	37	0.92
Cardiff	Regne Unit	35.25	58	15	51	15	21	12	61	1.21
Ipswich	Regne Unit	33.12	52	10	52	20	17	10	46	1.19
Norwich	Regne Unit	33.08	56	10	50	24	18	8	53	1.06
Reykjavik	Islàndia	33.94	50	12	35	23	24	18	68	1.51
Bergen	Noruega	33.5	49	10	39	17	36	8	53	1.21
Goteborg	Suècia	33.5	51	13	42	23	24	12	62	0.98
Umea	Suècia	34.2	50	14	50	25	21	4	39	1.26
Uppsala	Suècia	32.44	49	17	54	25	15	6	36	0.94
Wellington	Nova Zelanda	34.71	44	15	51	27	15	7	45	1.6
Christchurch	Nova Zelanda	34.04	52	13	51	23	17	8	40	1.56
Hawkes Bay	Nova Zelanda	35	61	18	61	24	11	4	47	1.5
Portland	Estats Units	35.12	53	13	58	22	10	10	40	1.45
Melbourne	Austràlia	34.57	48	20	50	24	15	11	35	1.27
Tartu	Estònia	31.7	59	8	47	19	28	7	56	0.75
Total (rang)		33.7 (31-37)	49.9 (44-63)	11.4 (7-20)	42.2 (29-61)	21.4 (11-29)	22.3 (10-36)	14.1 (4-31)	56.9 (35-84)	1.3 (0.7-2.9)

**Taula 4a. Percentatges de les possibles variables confusores individuals per centre**

Centre	País	Infeccions abans del 5 anys	IgE positiva a:				Rinitis	Verdet a casa en l'últim any
			Der p1	Gat	Herba timotea	Cladosporium		
South Antwerp	Bèlgica	8	22	9	16	3	27	20
Antwerp city	Bèlgica	15	28	10	15	7	27	26
Hamburg	Alemanya	8	19	11	25	5	25	20
Erfurt	Alemanya	12	15	8	20	4	12	9
Barcelona	Espanya	9	27	10	12	12	17	16
Galdakao	Espanya	7	19	3	9	1	14	12
Albacete	Espanya	10	7	2	10	0	25	10
Oviedo	Espanya	2	14	4	8	1	12	14
Huelva	Espanya	4	18	2	20	3	20	13
Grenoble	França	6	18	6	16	1	27	20
Dublin	Irlanda	10	34	8	16	2	24	19
Pavia	Itàlia	7	13	4	13	0	14	8
Turin	Itàlia	5	11	6	21	1	22	14
Verona	Itàlia	16	13	5	19	1	20	22
Groningen	Holanda	7	28	8	14	2	16	25
Bergen op Zoom	Holanda	9	24	6	19	5	15	13
Geleen	Holanda	7	27	9	21	4	20	20
Cambridge	Regne Unit	10	27	12	27	2	33	26
Cardiff	Regne Unit	11	26	8	13	2	23	24
Ipswich	Regne Unit	7	22	11	23	3	29	32
Norwich	Regne Unit	9	26	9	25	3	28	34
Reykjavik	Islàndia	12	9	7	11	7	24	6
Bergen	Noruega	12	13	8	16	1	18	4
Goteborg	Suècia	10	11	12	16	2	25	18
Umea	Suècia	13	6	11	14	3	30	15
Uppsala	Suècia	12	7	14	19	2	32	22
Wellington	Nova Zelanda	12	31	7	22	2	40	56
Christchurch	Nova Zelanda	12	33	15	33	3	35	52
Hawkes Bay	Nova Zelanda	11	31	9	25	1	36	57
Portland	Estats Units	7	19	13	34	5	41	46
Melbourne	Austràlia	15	32	9	29	2	47	55
Tartu	Estònia	15	10	5	9	7	17	21
Total (rang)		10.0 (2-16)	18.9 (6-34)	8.5 (2-15)	18.7 (9-33)	3.0 (0-12)	24.6 (12-47)	22.0 (4-57)

**Taula 4b. Percentatges de les possibles variables confusores individuals per centre**

## 6.2. Descripció de les variables a nivell de centre

De variables a nivell de centre en fem servir de dos tipus. Un són les variables que obtenim agregant variables individuals, com per exemple el percentatge d'individus de classe social baixa en un centre. Aquestes variables ja han estat descrites a la secció anterior, on les taules es presentaven estratificades per centre. L'altre tipus de variables són les que són variables purament de centre, que en aquest cas són només dues: el GNP per càpita del país i la temperatura mitjana anual del centre. A la **taula 5** es poden veure els valors d'aquestes variables.

Centre	País	GNP	Temperatura mitjana
South Antwerp	Bèlgica	24710	11.7
Antwerp city	Bèlgica	24710	11.7
Hamburg	Alemanya	27510	9.8
Erfurt	Alemanya	27510	7.8
Barcelona	Espanya	13580	16.5
Galdakao	Espanya	13580	14.9
Albacete	Espanya	13580	17.9
Oviedo	Espanya	13580	13.7
Huelva	Espanya	13580	20.8
Grenoble	França	24990	12.5
Dublin	Irlanda	14710	9.9
Pavia	Itàlia	19020	11.1
Turin	Itàlia	19020	14.7
Verona	Itàlia	19020	12.5
Groningen	Holanda	24000	8.7
Bergen op Zoom	Holanda	24000	9.8
Geleen	Holanda	24000	12.6
Cambridge	Regne Unit	18700	9.9
Cardiff	Regne Unit	18700	9.8
Ipswich	Regne Unit	18700	9.5
Norwich	Regne Unit	18700	9.4
Reykjavik	Islàndia	24950	5.1
Bergen	Noruega	31250	8.1
Goteborg	Suècia	23750	8.7
Umea	Suècia	23750	3.4
Uppsala	Suècia	23750	6.8
Wellington	Nova Zelanda	14340	12.4
Christchurch	Nova Zelanda	14340	11.8
Hawkes Bay	Nova Zelanda	14340	14.3
Portland	Estats Units	26980	12.6
Melbourne	Austràlia	18720	14.8
Tartu	Estònia	2860	5.5
Rang		2860-31250	3.4-20.8

**Taula 5. Valors de les variables a nivell de centre que no s'obtenen agregant variables individuals**



El GNP per càpita oscil·la entre els 2.860 \$ d'Estònia, que és amb diferència el centre més pobre de la mostra ja que la resta sobrepassen els 10.000 \$, fins als 31.250 \$ de Noruega. Com ja s'ha comentat anteriorment, el GNP per càpita té l'inconvenient que assigna un mateix valor a tots els centres d'un mateix país, i fins i tot quan només tenim un centre d'un país pot ser que aquest no estigui ben representat per aquest valor. Pel que fa a la temperatura mitjana anual, es pot observar la ja esperada diferència de climes. Mentre que a Umea tenen una mitjana de 3.4°, i la resta de països nòrdics tenen també una temperatura baixa, Huelva sobrepassa els 20° de mitjana. En aquesta variable és molt clar que la oscil·lació entre centres d'un mateix país és molt menor que la oscil·lació entre països.

## **6.3. Associacions crues de les variables individuals amb asma**

### **6.3.1. Variables socioeconòmiques**

La **taula 6** mostra les diferències que existeixen respecte a la prevalença d'asma en les diferents categories de les variables socioeconòmiques, sense ajustar-les per cap variable, ni tant sols per centre. S'observen diferències estadísticament significatives respecte a classe social, on els individus de la classe social més baixa o més poc benestant (classes IV-V) tenen una prevalença més alta d'asma que els de les classes socials més altes. També els que tenen un nivell d'estudis més baix són tenen una prevalença d'asma més alta que la resta. Els exposats a la feina a diferents contaminants tenen també una prevalença d'asma més alta en comparació amb els que no. Aquesta diferència és propera a la significació estadística.

Variable	Percentatge d'asmàtics	
Classe social I-II	7.4 %	} p=0.004
Classe social III no-manual	9.0 %	
Classe social III manual	7.7 %	
Classe social IV-V	10.2 %	
No classificable	9.4 % *	
Educació fins <16 anys (baixa)	10.5 %	} p=<0.001
Educació fins 16-19 anys (mitjana)	8.7 %	
Educació fins ≥20 anys (alta)	7.3 % *	
Exposició laboral a pols biològica, mineral, gasos o fums	9.7 %	} p=0.051
No exposició laboral a pols biològica, mineral, gasos o fums	8.2 % *	

**Taula 6. Diferències en les variables socioeconòmiques individuals entre asmàtics i no asmàtics**

### 6.3.2. Variables confusores

A la **taula 7** es mostren les prevalències d'asma en les diferents categories de les possibles variables confusores. Tal i com s'ha apuntat a la secció 3.4.2, aquestes variables s'han escollit perquè en alguna anàlisi prèvia havien tingut algun tipus d'associació amb l'asma.

En les nostres dades es pot observar que les dones tenen més asma que els homes en aquesta mostra, i que la diferència és estadísticament significativa. També hi ha diferències estadísticament significatives en els fills de pares asmàtics (un o els dos), que tenen una prevalença d'asma un 9% més alta que els fills de pares no asmàtics. Els qui han tingut infeccions respiratòries abans dels 5 anys també tenen més asma que els que no, igual com els grups de persones que tenen el test d'IgE positiu a qualsevol dels 4 al·lèrgens (*Der p 1*, gat, herba timotea i *cladosporium*), que també tenen una prevalença d'asma molt més gran que els no sensibilitzats (no al·lèrgics). Quan l'al·lèrgia es mesura per rinitis autoreportada el resultat és el mateix. La presència de verdet en alguna zona de la casa en l'últim any també comporta una prevalença d'asma més gran i estadísticament significativa. En canvi, no s'observen diferències significatives en les variables edat, tabaquisme, exposició passiva al tabac i nombre de germans.

Variable	Percentatge d'asmàtics	
≤ 30 anys	8.4 %	} p=0.427
30-39 anys	8.1 %	
≥ 40 anys	8.9 %	
Homes	7.4 %	} p<0.001
Dones	9.3 % *	
Història familiar d'asma	16.4 %	} p<0.001
No història familiar d'asma	7.3 % *	
No fumador	8.2 %	} p=0.638
Ex-fumador	9.0 %	
<20 cigarrets/dia	8.0 %	
≥20 cigarrets/dia	8.4 %	
Exposició passiva al tabac	8.4%	} p=0.870
No exposició passiva al tabac	8.3%	
0 germans	8.3 %	} p=0.958
1 germà	8.4 %	
2 o més germans	8.5 %	
Infeccions respiratòries abans dels 5 anys	13.3 %	} p<0.001
No infeccions respiratòries abans dels 5 anys	7.2 %	
No sap	17.0 %*	
IgE positiu a <i>Der p 1</i>	16.0 %	} p<0.001
IgE negatiu a <i>Der p 1</i>	6.6 % *	
IgE positiu a gat	22.6 %	} p<0.001
IgE negatiu a gat	7.0 % *	
IgE positiu a herba timotea	15.6 %	} p<0.001
IgE negatiu a herba timotea	6.7 % *	
IgE positiu a <i>cladosporium</i>	18.6 %	} p<0.001
IgE negatiu a <i>cladosporium</i>	8.0 % *	
Rinitis	17.7%	} p<0.001
No rinitis	5.3% *	
Verdet a casa (últims 12 mesos)	10.9%	} p<0.001
No verdet a casa (últims 12 mesos)	7.7 % *	

**Taula 7. Diferències en les variables confusores individuals entre asmàtics i no asmàtics**

## 6.4. Associacions crues de les possibles confusores i el nivell socioeconòmic

Perquè una variable sigui confusora d'una relació ha d'estar relacionada tant amb la resposta, asma en el nostre cas, com amb la variable explicativa en qüestió, que són en el nostre cas les variables socioeconòmiques. Conèixer quines relacions existeixen entre les possibles variables confusores i les variables socioeconòmiques ens permetrà doncs conèixer el "potencial" de cada variable a confondre la relació, però a més a més pot ajudar a l'hora d'interpretar i d'entendre el model.

La **taula 8** mostra els percentatges de les possibles variables confusores per les diferents categories de les variables socioeconòmiques i avalua si existeixen diferències estadísticament significatives. Podem veure-hi que hi ha més individus del grup d'edat més vell en les classes socials més altes, mentre que els joves són més freqüents a les classes socials baixes. Això podria explicar-se pel fet que els més joves no han pogut accedir encara als llocs de direcció i de responsabilitat perquè necessiten encara d'un temps d'aprenentatge i de poder "escalar" posicions dins de la seva empresa. Amb la variable educació passa tot el contrari: el grup d'edat més jove és més freqüent en el d'educació alta o mitjana, cosa que sembla raonable ja que les cohorts més antigues tenien menys possibilitats d'estudiar que les més actuals. La relació de l'edat amb l'exposició laboral a contaminants no és estadísticament significativa.

Quan mirem el percentatge de dones en les diferents categories de classe social podem observar que no segueixen un patró clar. Per una banda, les categories I-II i les IV-V estan bastant anivellades pel que fa a sexe. En canvi la III manual té una alt percentatge d'homes, cosa lògica si considerem que aquest tipus de feines han estat tradicionalment realitzades per homes. En canvi, la categoria III no manual té un percentatge més elevat de dones, i això és degut a que moltes de les ocupacions d'aquest grup són feines que tradicionalment han realitzat les dones, com per exemple, la de secretària. El sexe en l'educació queda desnivellat en favor dels homes, que tenen un percentatge més alt en el grup d'educació alta. En les feines que tenen una alta

exposició laboral a contaminants el percentatge de dones és sensiblement inferior, resultat també esperable ja que tradicionalment aquestes feines han estat més lligades als homes.

La història familiar d'asma no varia per classe social, però en canvi sí que ho fa per educació. El percentatge de persones que han tingut un dels pares amb asma és més gran en el grup de baixa educació. Pel que fa a l'exposició laboral, el fet que algun dels pares hagi estat asmàtic no hi té res a veure.

El tabaquisme mostra un patró molt marcat per classe social, de manera que els més fumadors són els de les classes socials més baixes. En l'educació el patró és més marcat encara, indicant per exemple que en els d'educació baixa hi ha el doble de fumadors de més d'un paquet diari que en els d'educació alta. La distribució respecte a exposició laboral mostra que els qui més fumen són els que estan exposats. Els patrons de tabaquisme passiu són molt semblants.

El nombre de germans també és una variable que varia segons el nivell socioeconòmic. Així, les famílies més nombroses es troben en les classes socials més baixes i en els nivells educatius més baixos, cosa que ja s'ha comprovat per altres fonts, com per exemple pel fet que els immigrants de països pobres tenen més natalitat que la gent autòctona del país ric. Els exposats laboralment a contaminants també tenen més germans, ja que aquestes feines estan fetes per gent de classe social més baixa.

Les infeccions respiratòries abans dels 5 anys d'edat no varien per classe social ni per educació, encara que en aquest cas està molt proper a la significació, i sí que existeix una relació amb exposició laboral a contaminants mitjançant la qual els exposats han tingut menys infeccions respiratòries de petits. Això podria ser degut al fet que les persones que han patit malalties respiratòries poden tendir a evitar aquest tipus de feines.

Les diferències en la sensibilització a al·lèrgens tendeixen a ser més grans per nivell d'educació que per classe social, on fins i tot en el cas que les diferències assoleixin la significació, els patrons no són clars. En canvi, per educació es pot observar que el grup de més baixa educació són els que estan menys sensibilitzats a al·lèrgè de gat, herba timotea i cladosporium. El fenomen que la sensibilització sigui més freqüent en nivells d'educació alts ha estat reportat per diversos estudis, i s'atribueix a exposicions ambientals diferents que tenen els individus de més alta educació i a actituds higièniques que adopten, com ara vacunacions, ús d'antibiòtics o una dieta diferent que poden fer que el sistema immunitari esdevingui més indefens a les al·lèrgies. No hi ha diferències en aquestes variables per l'exposició laboral a contaminants. La rinitis és més freqüent en les classes social més altes, en els nivells educatius alts, i en els no exposats a contaminants a la feina.

Finalment, la presència de verdet a casa és més freqüent en les cases de les classes socials altes, un resultat que no és l'esperat però que és estadísticament significatiu. Per educació i per exposició laboral no hi ha diferències.

Variables	Classe social					p	Educació				p	Exposició laboral a pols biològica, mineral, gasos o fums			p
	I-II	III-nm	III-m	IV-V	No Class.		Alta	Mitjana	Baixa	No		Sí			
≤ 30 anys	24.61	33.33	33.52	36.57	44.81	<0.001	32.43	34.9	16.39	<0.001	31.11	32.89	0.196		
30-39 anys	45.53	43.52	42.09	40.15	34.57		43.92	42.5	41.17		42.88	43.12			
≥ 40 anys	29.86	23.15	24.38	23.28	20.62		23.65	22.6	42.44		26.01	23.99			
Dones	49.49	66.81	25.86	52.38	69.86	<0.001	48.17	50.54	52.95	0.003	52.96	29.6	<0.001		
Història familiar d'asma	10.9	11.22	11.06	12.86	12.83	0.185	10.41	11.39	14.63	<0.001	11.28	12.38	0.219		
No fumador	46.57	44.5	34.59	36.99	46.83	<0.001	48.23	39.5	32.35	<0.001	43.06	36.38	<0.001		
Ex-fumador	23.62	21.19	20.63	19.67	16.38		21.68	21.51	20.54		21.83	18.95			
<20 cigarrets/dia	19.21	22.53	24.28	25.72	24.42		20.13	24.3	22.21		22.06	23.72			
≥20 cigarrets/dia	10.61	11.79	20.5	17.62	12.36		9.96	14.69	24.9		13.06	20.95			
Tabaquisme passiu	47.39	59.23	66.11	66.06	51.16	<0.001	48.8	60.64	69.4	<0.001	55.71	64.66	<0.001		
0 germans	41.49	38.91	35.16	35.06	42.04	<0.001	42.66	38.05	28.39	<0.001	39.5	33.2	<0.001		
1 germà	30.43	29.77	29.76	27.48	26.12		31.18	28.81	26.17		29.73	27.52			
2 o més germans	28.08	31.33	35.07	37.46	31.84		26.16	33.14	45.44		30.77	39.28			
Infeccions respiratòries abans dels 5 anys	10.46	10.18	9.13	9.75	10.97	0.546	10.9	9.61	8.86	0.085	10.35	8.02	0.010		
IgE positiu a <i>Der p 1</i>	19.83	17.68	20.41	17.73	15.77	0.009	18.39	20.04	16.98	0.014	18.79	19.85	0.341		
IgE positiu a gat	9.22	8.93	7.77	7.52	7.42	0.096	9.78	8.55	4.3	<0.001	8.59	7.81	0.326		
IgE positiu a herba timotea	19.83	18.28	17.69	17.62	18.7	0.167	20.31	19.14	11.95	<0.001	18.91	16.94	0.074		
IgE positiu a <i>cladosporium</i>	3.44	2.47	3.25	2.11	3.4	0.034	3.3	3.06	1.81	0.013	2.99	2.97	0.969		
Rinitis	27.34	25.01	21.3	21.15	26.74	<0.001	26.81	23.75	20.6	<0.001	25.08	21.45	0.003		
Verdet a casa (últims 12 mesos)	23.62	22.1	19.53	21.02	22.1	0.004	21.48	22.63	21.21	0.305	22.15	20.68	0.207		

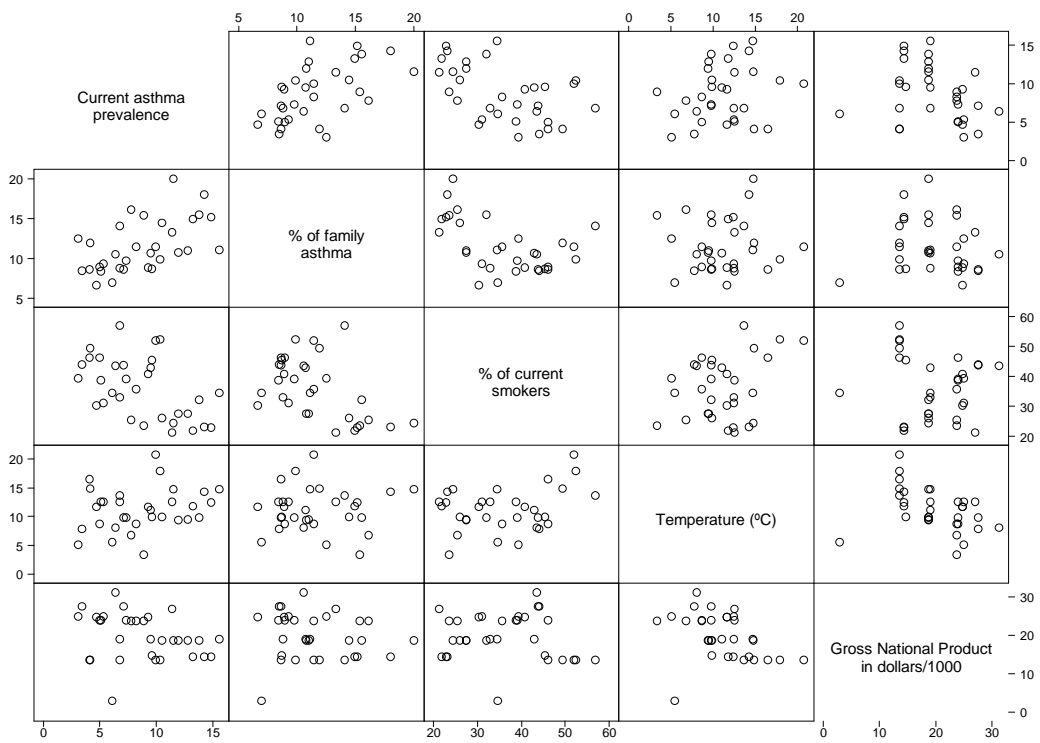
**Taula 8. Percentatges de les possibles variables confusores per les diferents categories de les variables socioeconòmiques.**

## 6.5. Associacions de les variables a nivell de centre i asma

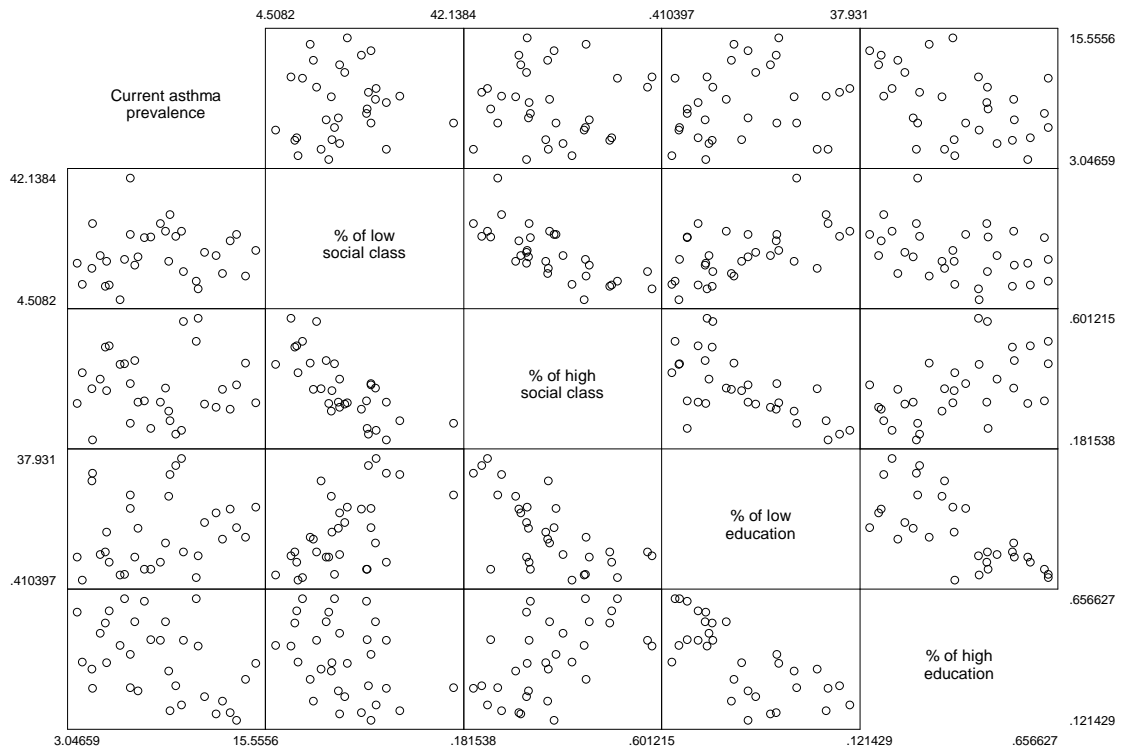
Tot i que els models multinivell valoren la relació que té la variable a nivell de grup amb el risc individual, en aquest capítol exploratori creuarem les diferents variables explicatives a nivell de centre amb el percentatge (o prevalença) d'asmàtics del centre.

A les **figures 1-2** es poden observar els gràfics entre la prevalença d'asma dels centres i les altres variables a nivell de centre. Per ajudar a valorar la correlació a nivell de centre que veiem visualment als gràfics, la **taula 9** mostra les correlacions d'Spearman d'aquestes relacions. La justificació de l'ús de la correlació d'Spearman es pot trobar a la secció 5.1.1. Les correlacions que són estadísticament diferents de zero amb la prevalença d'asma són les de prevalença d'història familiar d'asma, percentatge de fumadors i percentatge d'individus d'alta educació. Aquestes són també les variables que en els gràfics semblen més relacionades amb la prevalença d'asma. Dels gràfics i de la taula també en podem destacar les correlacions existents entre el percentatge d'asma familiar i el percentatge de fumadors, que segurament ens indica el mateix que el gràfic d'asma en l'actualitat contra el percentatge de fumadors, és a dir, que els centres amb més asma són també els centres que tenen més fumadors. A més a més, la temperatura del centre té una correlació alta amb el percentatge de baixa educació, i aquesta última variable té una correlació negativa important amb el GNP. També, òbviament, les variables d'educació i classe social tenen una alta correlació entre elles.





**Figura 1. Matrix plot de la prevalença d'asma per centre contra les variables: prevalença d'asma familiar, percentatge de fumadors, temperatura mitjana anual i GNP per càpita**



**Figura 2. Matrix plot de la prevalença d'asma per centre contra variables de percentatge de baixa/alta educació i alta/baixa classe social**

	Prevalença d'asma	% asma familiar	% fumadors	Temperatura	GNP	% Baixa classe social	% Alta classe social	% Baixa educació	% Alta educació
Prevalença d'asma	1,00								
% asma familiar	0,58	1,00							
% fumadors	-0,53	-0,49	1,00						
Temperatura	0,23	0,08	0,16	1,00					
GNP	-0,33	-0,26	-0,11	-0,44	1,00				
% Baixa classe social	0,13	0,11	0,25	0,17	-0,25	1,00			
% Alta classe social	-0,08	-0,00	-0,34	-0,10	0,36	-0,73	1,00		
% Baixa educació	0,25	0,02	0,39	0,62	-0,64	0,55	-0,64	1,00	
% Alta educació	-0,46	-0,12	-0,03	-0,34	0,66	-0,37	0,57	-0,77	1,00

**Taula 9. Correlacions d'Spearman entre les variables a nivell de centre.**

# Capítol 7. Selecció de models

## 7.1. Descripció

Seguint l'estratègia de modelització explicada a la secció 5.1.2.1, modelitzem primer les variables a nivell individual. Així, fem models de regressió logística amb centre com a variable que defineix el segon nivell i amb efectes aleatoris a la constant i, si són significatius, efectes aleatoris en els pendents, però no introduïm de moment variables a nivell de centre al model. Fent-ho arribem a tres models que són pràcticament equivalents pel que fa a la qualitat de l'ajust, que són els que es mostren a la **taula 10**. Així, de les variables socioeconòmiques i de les possibles variables confusores de la **taula 7**, només les de la **taula 10** són les que romanen al model.

Les diferències entre aquests tres models són degudes a que de les tres variables socioeconòmiques, classe social, educació, i exposició laboral a pols biològica, mineral, gasos o fums, només dues poden estar simultàniament al model, probablement degut a la correlació que existeix entre elles (**taula 3**). L'efecte de la resta de variables és bàsicament el mateix en els tres models.

Cap dels odds ratios de la **taula 10** mostren una heterogeneïtat significativa, però la constant sí que continua tenint efectes aleatoris significatius malgrat la inclusió de les variables a nivell individual. Per tal d'intentar explicar part d'aquesta variabilitat entre centres que encara queda inexplicada, s'introdueixen al model variables a nivell de centre. De totes les variables proposades a les **figures 1-2**, només les variables temperatura mitjana anual ( $p=0.10$ ), GNP per càpita ( $p=0.07$ ), i percentatge de d'individus d'educació alta ( $p=0.003$ ) resulten significatives si s'inclouen al model (a qualsevol dels 3 models de la **taula 10**), és a dir, només aquestes variables tenen un efecte un cop s'ha ajustat per les variables individuals. Si ajustem les variables a nivell

de centre l'una per l'altra, només el percentatge d'individus amb educació alta es manté significatiu i la resta de relacions a nivell de centre desapareixen.

	Model 1	Model 2	Model 3
Efectes fixos	Odds Ratio (Interval de confiança al 95%)		
Nivell d'individu			
Edat (referència: <30 anys)			
30-39 anys	0.969 (0.816, 1.152)	0.979 (0.824, 1.164)	0.944 (0.794, 1.121)
>40 anys	1.146 (0.942, 1.394)	1.184 (0.977, 1.435)	1.106 (0.909, 1.346)
Sexe: dona	1.256 (1.085, 1.455)	1.278 (1.096, 1.489)	1.319 (1.137, 1.531)
Història familiar d'asma	2.028 (1.707, 2.410)	2.042 (1.705, 2.446)	2.042 (1.705, 2.446)
Tabac (referència=no fumadors)			
Ex fumador	1.192 (1.011, 1.406)	1.203 (0.995, 1.455)	1.191 (0.985, 1.441)
Fumador: < 20 cigs./dia	1.148 (0.946, 1.394)	1.162 (0.957, 1.411)	1.157 (0.953, 1.405)
Fumador: ≥ 20 cigs./dia	1.258 (0.927, 1.708)	1.288 (1.026, 1.617)	1.260 (1.002, 1.585)
Infeccions respiratòries abans dels 5 anys d'edat	1.812 (1.478, 2.221)	1.808 (1.474, 2.216)	1.815 (1.480, 2.225)
IgE positiva a Der p 1	1.696 (1.359, 2.116)	1.689 (1.424, 2.002)	1.694 (1.428, 2.009)
IgE positiva a <i>Fel d 1</i>	2.209 (1.719, 2.838)	2.201 (1.795, 2.699)	2.206 (1.802, 2.699)
IgE positiva a <i>Cladosporium</i>	1.517 (1.262, 1.824)	1.516 (1.095, 2.099)	1.489 (1.075, 2.061)
Rinitis	2.808 (2.331, 3.382)	2.798 (2.402, 3.260)	2.784 (2.390, 3.244)
Classe social: referència = I-II			
III no-manual*	1.230 (0.986, 1.535)	1.269 (1.039, 1.549)	
III manual	1.154 (0.996, 1.337)	1.141 (0.824, 1.580)	
IV-V*	1.499 (1.232, 1.823)	1.499 (1.083, 2.076)	
No classificable	1.251 (0.999, 1.568)	1.306 (0.953, 1.791)	
Educació: referència = Alta			
Mitjana	0.877 (0.707, 1.088)		0.853 (0.686, 1.060)
Baixa	0.793 (0.624, 1.007)		0.732 (0.579, 0.926)
Alta exposició laboral a pols biològica o mineral, gas o fums		1.224 (0.971, 1.542)	1.275 (1.034, 1.573)
Efectes aleatoris	Variància (error estàndard de la variància)		
$\sigma_u^2$ (centre) ‡	0.116 (0.041)	0.129 (0.044)	0.116 (0.041)

† Variables incloses per una  $p < 0.10$  entre les variables de la taula 5. Les variables que finalment no romanen al model són: exposició passiva al tabac, nombre de germans, IgE positiva a herba timotea i verdet a casa durant l'últim any.

‡ Variància deguda al centre

\* I = professionals; II = ocupacions intermèdies; III no-manual = treballadors qualificats en ocupació no manual; III manual = treballadors qualificats en ocupació manual; IV = treballadors parcialment qualificats en una ocupació no manual; V = treballadors no qualificats en una ocupació manual.

**Taula 10.** Models† de regressió logística per la variable "asma a l'actualitat". La variable "centre" és tractada com un efecte aleatori. En aquests models no s'han inclòs variables a nivell de centre

Com que els tres models de la **taula 10** són pràcticament equivalents i els resultats són molt semblants quan es fa servir qualsevol dels tres models, prenem, per exemple, el model 1. El model final, amb variables a nivell de centre, és el de la **taula 11**.

La variància deguda al centre quan no incloem cap variable al model (pas 1 de l'estratègia de modelització descrita al capítol 6) és 0.177. Quan incloem només variables a nivell d'individu (model 1 de la **taula 10**) la variància deguda al centre és un 34% menor que en el model sense variables. Quan a més incloem la variable a nivell de centre "percentatge d'alta educació", la variància deguda al centre és un 55% menor que en el model sense variables explicatives.

	Odds Ratio (IC al 95%)
Efectes fixos	
Nivell d'individu	
Edat (referència: <30 anys)	
30-39 anys	0.972 (0.818, 1.155)
>40 anys	1.158 (0.952, 1.409)
Sexe: dona	1.253 (1.075, 1.460)
Història familiar d'asma	2.037 (1.701, 2.440)
Tabac: referència = no fumadors	
Ex fumador	1.196 (0.989, 1.447)
Fumador: < 20 cigs./dia	1.150 (0.947, 1.396)
Fumador: ≥ 20 cigs./dia	1.264 (1.005, 1.590)
Infeccions respiratòries abans Dels 5 anys d'edat	1.807 (1.473, 2.215)
IgE positiva a Der p 1	1.671 (1.409, 1.982)
IgE positiva a <i>Fel d 1</i>	2.230 (1.819, 2.734)
IgE positiva a <i>Cladosporium</i>	1.518 (1.097, 2.102)
Rinitis	2.828 (2.427, 3.295)
Classe social: referència = I-II	
III no-manual*	1.236 (1.006, 1.518)
III manual	1.157 (0.927, 1.443)
IV-V*	1.503 (1.199, 1.883)
No classificable	1.239 (0.900, 1.705)
Educació: referència=Baixa	
Mitjana	0.897 (0.719, 1.120)
Alta	0.841 (0.656, 1.079)
Nivell de centre	
% d'alta educació	0.988 (0.981, 0.996)
Efectes aleatoris	
	Variància (error estàndard de la variància)
$\sigma_u^2$ (centre) ‡	0.080 (0.031)

† Variables incloses per una  $p < 0.10$  entre les variables de la taula 5. Les variables que finalment no romanen al model són: exposició passiva al tabac, nombre de germans, IgE positiva a herba timotea i verdet a casa durant l'últim any.

‡ Variància deguda al centre

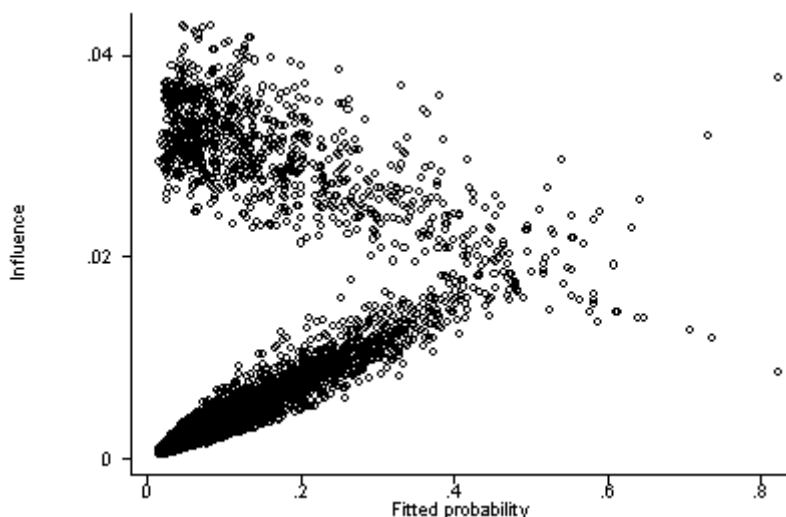
\* I = professionals; II = ocupacions intermèdies; III no-manual = treballadors qualificats en ocupació no manual; III manual = treballadors qualificats en ocupació manual; IV = treballadors parcialment qualificats en un ocupació no manual; V = treballadors no qualificats en una ocupació manual.

**Taula 11:** Associació multivariant † (Odds Ratio (IC 95%)) entre "asma a l'actualitat i variables a nivell individual i de centre fent servir un model de regressió logística multinivell.

## 7.2. Validació

### 7. 2. 1 Residus a nivell d'individu

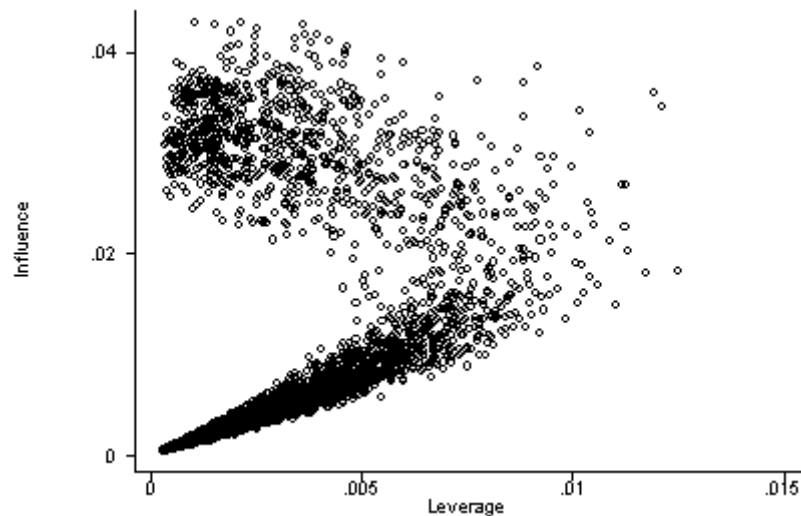
Els residus a nivell 1 els tractem com els residus d'un model logístic qualsevol. Així, tal i com s'ha descrit a la secció 5.2.1, el gràfic més important en el que ens hem de fixar és el de la influència contra els valors previstos. Aquest gràfic el podem trobar a la **figura 3**.



**Figura 3.** Gràfic de la influència contra la probabilitat predita

El primer que es pot observar en aquesta figura és que no hi ha uns valors d'influència molt separats de la resta. Els més alts els trobem quan la probabilitat estimada és menor de 0.1, justament un dels llocs on haurien de ser més petits segons el comportament esperat descrit a la secció 5.2. Podem agafar com a observacions més influents, per exemple, les que tenen un valor d'influència més gran que 0.04. Són 19 observacions. D'aquests, no hi ha cap patró covariable que es repeteixi. Tots són asmàtics, i no tenen la majoria dels factors de risc d'asma que ens diu el model: 18 d'ells no tenen sensibilització al gat, 16 no tenen rinitis, 16 no tenen història familiar d'asma, 16 no tenen sensibilització a cladosporium, etc. Si excloem aquestes observacions del model, els canvis en els valors estimats són mínims.

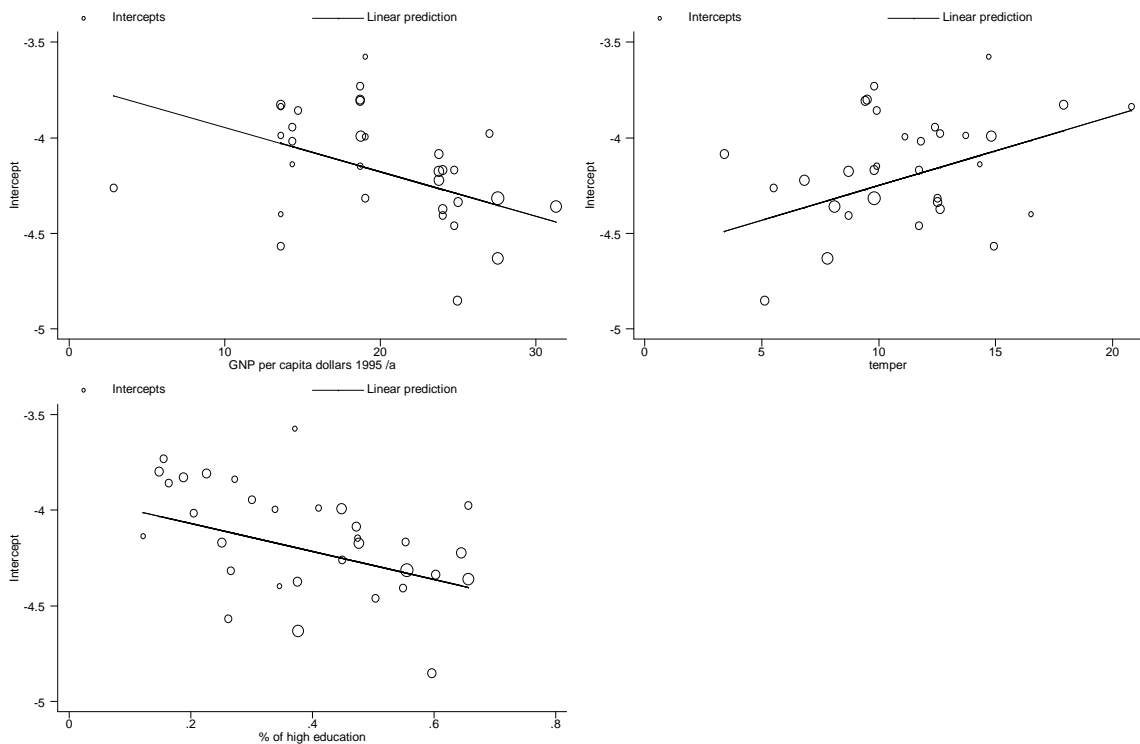
També s'ha explicat a la secció 5.2.1 que el gràfic de la influència contra el leverage ens pot ser útil. A la **figura 4** hi tenim aquest gràfic. S'hi observa que les observacions amb més influència no són les que tenen més leverage. Això ens indica que aquestes observacions són influents per tenir un residu gran, i no a causa d'un leverage gran. Com que aquestes observacions estan en una probabilitat prevista menor de 0.10, el leverage no és interpretable com una distància de les x's a la seva mitjana, tal i com s'ha explicat a la secció 5.2.1, però en tot cas, ja hem vist que la seva influència no és deguda a un leverage elevat.



**Figura 4.** Gràfic de la influència contra el leverage

## 7. 2. 2 Residus a nivell de centre

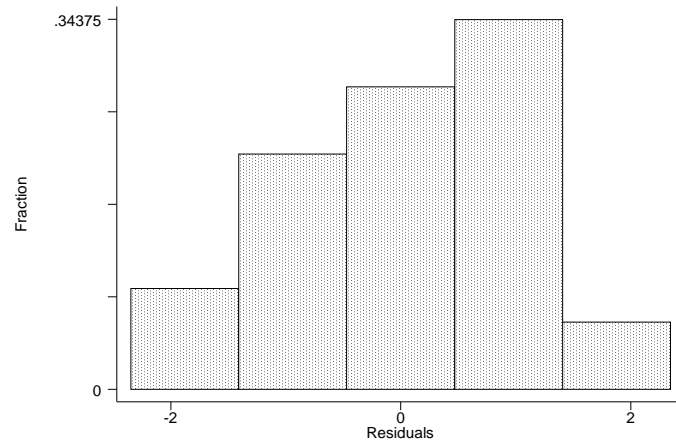
El primer que es pot fer a nivell de centre és veure si les variables a nivell de centre tenen realment una relació lineal amb les constants del model sense variables a nivell de centre, tal i com s'ha explicat a la secció 5.2.2. La **figura 5** mostra els gràfics de les constants dels diferents centres (obtingudes a partir del model 1 de la **taula 10**) contra les variables a nivell de centre que ens han sortit associades amb la resposta. Observant la **figura 5**, podem acceptar que la relació és lineal i que no és deguda a uns pocs centres.



**Figura 5.** Plots de les constants de cada centre (obtingudes del model 1 a la taula 10) contra variables a nivell de centre. La recta obtinguda per mínims quadrats ponderats hi està sobreposada. Els símbols són proporcionals als pesos (nombre d'individus en cada centre)

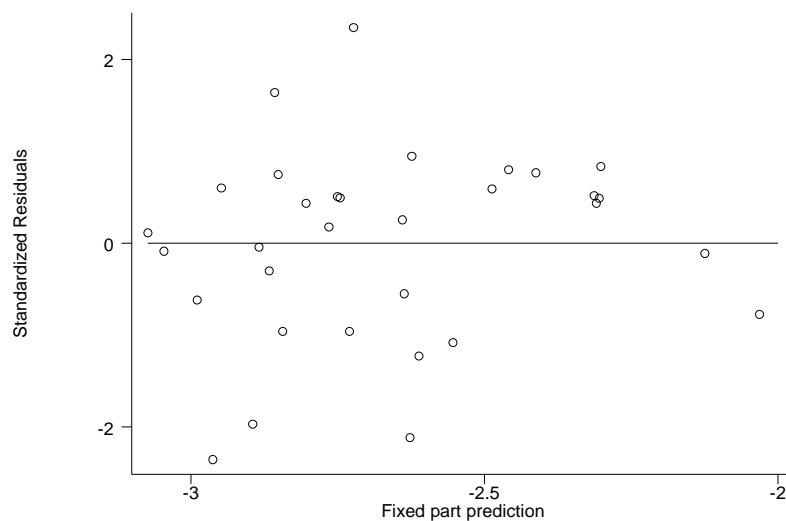
Un cop validat que les variables a nivell de centre estan ben entrades com a variables contínues, podem entrar a fer l'anàlisi de residus. Podem començar valorant la normalitat dels residus fent-ne un histograma (**figura 6**). El gràfic mostra una asimetria aplanada per l'esquerra i per tant s'allunya de la normalitat. Cal tenir en compte, però, que només són 32 punts.



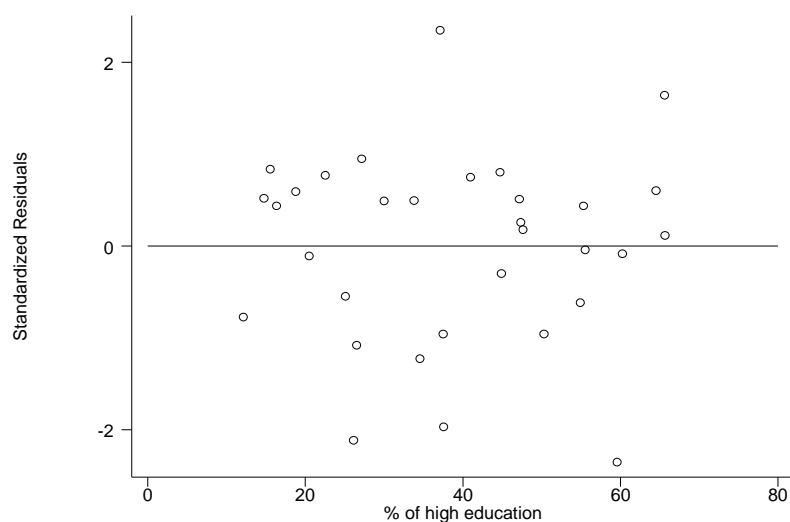


**Figura 6.** Histograma dels residus estandarditzats a nivell de centre

També podem fer el gràfic de residus estandarditzats contra valors previstos, que el trobem a la **figura 7**. Els valors previstos, com s'ha explicat a la secció 5.5.2., són les mitjanes per centre de la part fixa del model. En aquest gràfic hi veiem que pels valors previstos més alts la variància és menor, i a més sembla que els residus no estan centrats a zero en aquesta zona. No obstant, en aquest tram és on hi ha menys observacions. Els dos punts amb el residu més gran i positiu semblen estar separats de la resta, i poden tenir gran part de culpa en el fet que els residus no estiguin centrats a zero. Aquests dos punts són Turin i Portland. A la **figura 8**, els residus estan creuats amb l'única variable a nivell de centre del model, el percentatge d'alta educació, i hi podem veure novament els dos punts allunyats de la resta. Mentre que Portland està en un extrem, Turin està al centre de la variable explicativa. En els propers gràfics s'analitza la influència i es veurà quin efecte poden tenir aquests dos punts.



**Figura 7.** Residus estandarditzats a nivell de centre contra els valors previstos

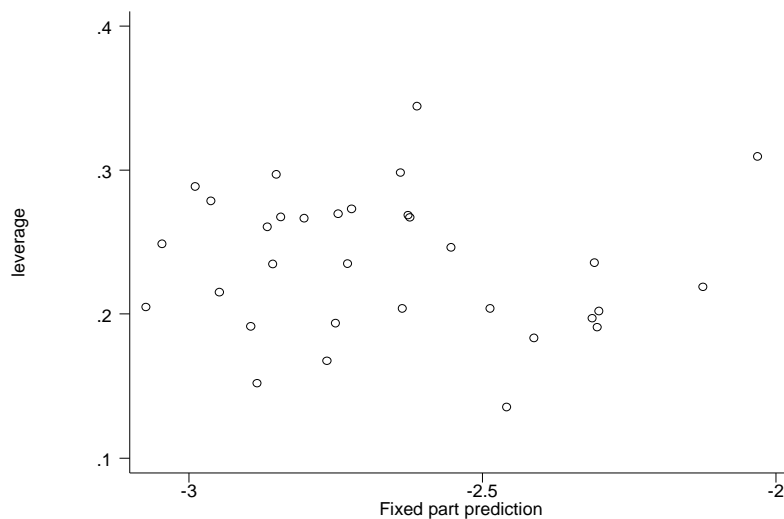


**Figura 8.** Residus estandarditzats a nivell de centre contra el percentatge d'alta educació

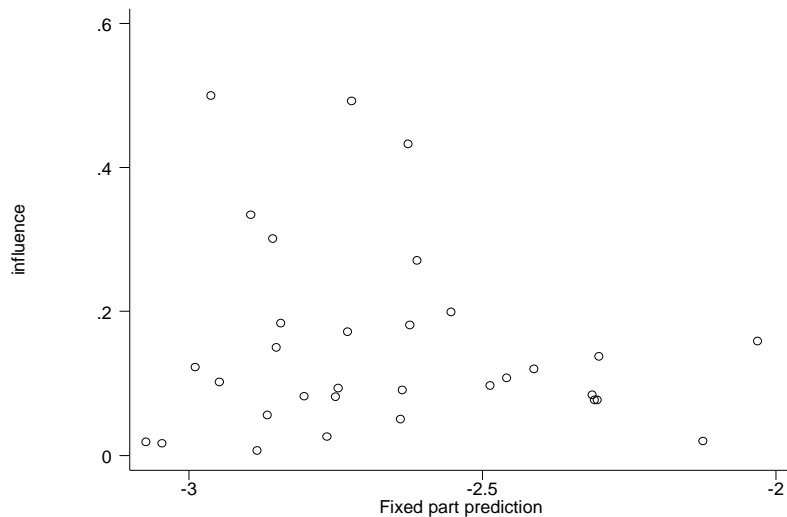
Comencem doncs a analitzar altres diagnòstics. A la **figura 9** hi tenim el leverage a nivell de centre contra el valor previst. S'observen dos punts que tenen un leverage més alt que la resta. El que té el leverage més alt és Barcelona. Aquest centre té una de les n's més petites de l'estudi, i a més està al 20% inferior en les variables GNP, percentatge de dones, percentatge d'història familiar d'asma i al 10% superior en les variables percentatge de fumadors i fumadors passius, percentatge de baixa educació i percentatge de sensibilitzats a cladosporium. El segon pel que fa a leverage és Hawkes Bay. Aquest centre té la n més petita de totes, és també el que té un percentatge més baix d'alta educació, el que té un percentatge més alt de no fumadors i de verdet a casa, i un dels que té un percentatge més alt de dones, d'història familiar d'asma i de sensibilització a Der p 1.

A la **figura 10** tenim la influència a nivell de centre contra els valors previstos. Hi ha tres punts que s'allunyen de la resta, que són Reykjavik i Turin (les dos amb més influència) i Galdakao. Reykjavik és el centre que té menys asma, amb una prevalença del 2.8%, i Turin, en canvi, és el que la té més alta, un 15.72%. A més, Reykjavik és un dels centres que té més observacions, mentre que Turin és el segon que en té menys. Reykjavik és el centre que té el residu més baix a nivell 2 i Turin és el que el té més alt. Galdakao també és dels centres amb menys asma, i té un dels residus més negatius a

nivell 2. Pel que fa a la variable a nivell de centre que mantenim al model final, el percentatge d'alta educació, Turin té un valor que està a la mediana dels centres, el de Galdakao tampoc destaca i el de Reykjavik sí que està en el 15% superior. A la **figura 5** podíem veure com, en el gràfic de les constants del model contra el percentatge d'educació, Reykjavik és el punt que està més avall en l'eix de les y, i dels que està més apartat del centre en l'eix de les x. Per tant, aquest punt té una influència considerable en la recta estimada en aquest gràfic, i en conseqüència, té bastant influència en l'efecte de la variable "percentatge d'alta educació". Si estimem el model sense incloure el centre Reykjavik, el model pateix lleus i poc importants canvis en les estimacions per les variables a nivell d'individu, però l'odds ratio de la variable percentatge d'alta educació al centre continua sent el mateix amb la mateixa significació estadística. Per tant, doncs, la relació trobada no és deguda al centre Reykjavik. L'efecte de Turin, per les raons abans comentades, és molt inferior.



**Figura 9.** Gràfic del leverage a nivell 2 contra els valors previstos



**Figura 10.** Gràfic de la influència a nivell 2 contra els valors previstos

## 7.3. Interpretació

Prenent com a model final el de la **taula 11** podem procedir a intentar interpretar-lo. Com ja s'ha explicat a la secció 5.3, els efectes de cada variable són per quan la resta de variables del model romanen constants. Seguidament es comenten un per un els efectes de totes les variables.

### 7.3.1. Variables socioeconòmiques

#### 7.3.1.1. Nivell d'individu

- Pel que fa a la classe social, constatem que els que tenen menys asma són els de la categoria de referència, els de classe social més alta. La categoria III no manual (treballadors qualificats en ocupació no manual) i la categoria IV-V (treballadors parcialment qualificats i treballadors no qualificats) tenen un risc estadísticament més alt que la categoria de referència. El risc de la categoria III manual també és més alt i està al límit de la significació.

L'increment de risc en la classe social més baixa respecte a la més alta és d'un 50%. El risc entre els no classificables també és més elevat, encara que no acaba de ser estadísticament significatiu.

- La variable educació també indica que els qui tenen més asma són els de nivell més baix, malgrat que l'educació a nivell individual deixa de ser significativa quan el model conté la variable educació a nivell de centre.
- Pel que fa a estar exposat a pols mineral o biològica, gas o fums a la feina, suposa incrementar el risc d'asma en un 25%, encara que en el model de la **taula 11** no està inclosa perquè només dues de les tres variables socioeconòmiques poden estar al model a l'hora.

### 7.3.1.2. Nivell de centre

- El percentatge de persones del centre amb alt nivell d'educació és estadísticament significatiu en el model de la **taula 11**, indicant que pertànyer a un centre amb un alt percentatge d'alta educació és protector de cara a tenir asma (per cada 1% més de persones amb educació alta que té el centre, el risc és un 12% menor). Quan aquesta variable és inclosa a nivell de centre, alta educació a nivell individual perd la significació ( $p=0.17$ ), indicant una importància més alta de la variable educació a nivell de centre.
- Incloure la variable GNP com a única variable a nivell de centre en el model també provoca una reducció significativa de la variabilitat deguda al centre. No obstant, quan també incloem la variable percentatge d'alta educació el GNP perd la significació, mentre que el percentatge d'alta educació es manté significativa. Això pot ser degut a la correlació alta que existeix entre les dues ( $r=0.66$ ) i també a que tenim un nombre relativament petit de centres.

### 7.3.2. Variables confusores

En les dones, el risc és al voltant d'un 30% més gran que en el homes. Els que tenen història familiar d'asma tenen més del doble de possibilitats de tenir asma que els que no. Els ex-fumadors també tenen més risc que els no fumadors, tot i que aquesta variable està al límit de la significació. Els fumadors de menys de 20 cigarretes semblen tenir menys risc que els ex-fumadors, però els grans fumadors (més d'un paquet al dia) són els que tenen el percentatge més alt d'asmàtics (un 25% més que els no fumadors). Haver tingut infeccions respiratòries abans dels 5 anys també es mostra com un factor de risc per desenvolupar asma. Pel que fa a la sensibilització a al·lèrgens, els sensibilitzats a *Der p 1* tenen un 70% més de risc, els que ho estan al gat un 120% i els que ho són a *Cladosporium* al voltant del 50% més de risc. La variable rinitis és la que mostra una associació més gran amb asma (OR=2.8).

A nivell de centre, de les variables de la secció 3.5.1, només la variable temperatura anual del centre té una relació amb la resposta, indicant que els que viuen en un centre amb una temperatura més alta tenen més risc de patir asma. No obstant, quan també introduïm al model la variable percentatge d'alta educació, l'associació amb la temperatura desapareix. La correlació entre aquestes dues variables és moderada ( $r = -0.34$ ). Cal tenir en compte que tenim molt pocs centres per buscar associacions amb més d'una variable.

Cap de les beta (o dels odds ratio), a part de la constant del model, presenta una heterogeneïtat significativa entre centres, i per tant les relacions són homogènies per centre. Així doncs, només tenim un efecte aleatori que afecta a la constant dels models, és a dir, a l'odds d'asma en l'últim any per la categoria de referència (edat mitjana, home, sense història familiar d'asma, no fumador, sense haver tingut infeccions respiratòries abans dels 5 anys, no sensibilitzat ni a *Der p 1* ni a gat ni a *Cladosporium*, sense rinitis, de classe social alta, d'educació baixa i sense exposició laboral a compostos químics).

# Capítol 8. Conclusions

## 8.1. Discussió de les troballes

En aquesta secció es presenten els resultats que trobem més enllà del significat numèric que s'ha donat als capítols 6 i 7, comparant els resultats amb altres estudis, intentant interpretar les relacions i el per què poden estar produïdes, i analitzant les possibles limitacions que té l'estudi.

### 8.1.1. Resum dels principals resultats

Totes les variables socioeconòmiques individuals indiquen una prevalença d'asma més alta en les classes menys afavorides, que són: les classes social més baixes (IV i V: treballadors parcialment qualificats en una ocupació manual i treballadors no qualificats), els que han deixat els estudis abans dels 16 anys i aquells que tenen una alta exposició laboral a pols biològica o mineral, gasos o fums. A nivell de centre, els individus que pertanyen a un centre amb un baix percentatge d'alta educació tenen un risc més alt de patir asma, independentment de l'estatus socioeconòmic individual. Així, un baix nivell socioeconòmic del centre té un efecte afegit en el risc de patir asma, i a més aconsegueix explicar part de la variància entre centres en la prevalença d'asma. A més, totes les associacions trobades han estat estadísticament homogènies. És a dir, els odds ratio globals que trobem no varien per centre, o dit d'una altra manera, podem considerar que l'associació és la mateixa en tots els centres. Aquesta conclusió s'extreu del fet que cap dels pendents ha tingut un efecte aleatori associat estadísticament diferent de zero.

### **8.1.2. Aportacions de l'anàlisi multinivell**

El fet d'usar models multinivell ha permès avaluar l'efecte de les variables socioeconòmiques a nivell d'individu i a nivell de centre simultàniament, ajustades les unes per les altres. A més, contràriament als estudis ecològics, que tindrien com a resposta la prevalença d'asma del centre i per tant l'associació trobada només es podria interpretar a nivell de centre, l'efecte de les variables a nivell de centre en els models multinivell està lligat a la resposta individual i no a la prevalença del centre. El percentatge d'alta educació del centre redueix la variància entre centres no explicada un 31%. Quan incloem la variable a nivell de centre en el model, els odds ratio de la variable educació s'atenuen (el coeficient d'alta educació es redueix un 25%, perdent la significació estadística). Existeix, doncs, confusió entre la variable a nivell de centre i la variable a nivell individual. No obstant, la classe social a nivell individual manté la significació estadística quan s'ajusta per educació a nivell de centre.

### **8.1.3. Consistència amb altres estudis**

Els resultats obtinguts són consistents amb la majoria d'altres estudis. Així, (Court 2002, Bozicevic 2000, Lewis 2001, Litonjua 1999, Pallasaho 1999, Isoaho 1994, Sibbald 1992) també troben que les classes socials més baixes pateixen més asma, encara que Chen (2002) i Montnemaury (2001) no troben associació.

Quan l'educació és la mesura de nivell socioeconòmic, un estudi troba com nosaltres una relació negativa entre alta educació i asma (Litonjua 1999). A nivell de centre, tots els indicadors socioeconòmics mostren que estar en un centre amb una mitjana de nivell socioeconòmic alt redueix la prevalença d'asma, independentment de l'estatus social individual. Aquest fet no contradiu l'evidència que en les zones rurals i de subsistència l'asma és baix, (Odhiambo 1998, Yemaneberhan1997), perquè els centres participants en el nostre estudi pertanyen tots a països desenvolupats. Stewart (2001) va trobar una relació positiva entre GNP i asma en nens, en la direcció contrària a la nostra, però el seu estudi incloïa un rang més gran de GNP, incloent països subdesenvolupats. Així, cal remarcar que les nostres conclusions són només aplicables a



països amb un GNP que estigui en el rang dels nostres centres (2.860-31.250 en dòlars de 1995, malgrat que Estònia té un GNP molt inferior a la resta, i el segon país amb GNP més petit té 13.580) i amb accés universal als serveis de salut. Saldmon (1999), Litonjua (1999) i Erzen (1997) troben un increment en la prevalença d'asma en adults en les àrees més deprimides d'Austràlia i dels Estats Units, en concordància amb els nostres resultats. No obstant, les nostres dades no permeten diferenciar petites àrees dins d'una mateixa ciutat, cosa que impedeix examinar les àrees més pobres de les ciutats, on s'han trobat índex d'asma desproporcionadament alts (Weiss 1992).

### **8.1.4. Interpretació de les mesures socioeconòmiques**

Un aspecte important dels nostres models és la inclusió de diverses mesures de nivell socioeconòmic en el mateix model, i és important saber les diferències que hi ha entre elles. L'educació és una mesura d'accés a la informació sobre aspectes que poden promoure o deteriorar la salut, mentre que la classe social mesura una barreja entre el que mira l'educació, el nivell de benestar material, i altres aspectes relacionats amb la ocupació particular, com el prestigi, el poder, i possibles exposicions ocupacionals. (Álvarez-Dardet 1995). A més a més, l'educació és una mesura estable en la vida d'una persona, és dir, que en la majoria dels casos no hi ha canvis en la categoria d'una persona una vegada ha acabat els seus estudis, mentre que els canvis en la ocupació sí que poden implicar un canvi de classe social. En el nostre estudi hi ha una correlació moderada-alta entre classe social, educació i alta exposició laboral a contaminants, però aquestes variables mantenen efectes independents en la resposta ja que posant-ne dues al model, les dues romanen estadísticament significatives. Aquest fet reforça que la classe social, mesurada a través de l'ocupació, i el nivell d'educació són mesures de nivell socioeconòmic diferents que mesuren coses diferents, i també que l'efecte de la classe social no és degut exclusivament a l'exposició ocupacional a contaminants.

Pel que fa a les variables a nivell de centre, el GNP per càpita podria ser una aproximació de la situació de la sanitat en un país (Stewart 2001), però aquesta mesura no ens permet diferenciar entre centres del mateix país. Per contra, el percentatge d'alta educació en el centre sí que pot fer-ho, però té el defecte que està basat en la mostra en

lloc de en tota la població, afegint una component d'error mostral a la mesura (Diez-Roux 1997).

### **8.1.5. Possibles explicacions a l'associació**

Molts dels factors de risc per l'asma estan relacionats amb el nivell socioeconòmic i s'han suggerit com a la possible explicació de la relació. En el nostre estudi se n'han tingut en compte molts d'ells (condició atòpica, tabaquisme actiu i passiu, edat, sexe, grandària de la família, infeccions a la infància, exposicions laborals i haver tingut verdet a casa durant l'últim any), però la diferència d'asma per nivell socioeconòmic encara perdura. Altres factors que s'han suggerit (Volmer 2001, Isoaho 1994, Rona 2000) i que no s'han tingut en compte en la nostra anàlisi són: contaminació ambiental de l'aire, obesitat, inactivitat, dieta, baix pes al néixer i altres variables de la primera fase de la vida, tipus de combustible que es fa servir a casa i males condicions de l'habitatge.

Les diferències en la prevalença d'asma per classe social s'havien atribuït a diferències en l'exposició a al·lèrgens (Chen 2002, Peat 96), però nosaltres trobem una diferència socioeconòmica fins i tot després d'ajustar per la sensibilització (IgE positiva) als al·lèrgens més freqüents a les nostres zones. Això suggereix que altres exposicions relacionades amb la pobresa, a part dels al·lèrgens, són els responsables de les diferències en la prevalença d'asma. Lewis (1995) va arribar a la mateixa conclusió quan tractava d'explicar les diferències trobades en les zones urbanes més pobres dels Estats Units, i Wieringa (1997) va afirmar que les diferències en la sensibilització als àcars de la pols explicaven l'aparició d'asma en la infància però no els símptomes en adults quan comparava el centre urbà amb suburbis a Bèlgica. En un altre article, Wieringa (98) trobava diferències en els símptomes d'asma entre el centre urbà i els afores en adults però no en nens, i suggeria que les diferències podien ser degudes a un efecte progressiu i a llarg termini de l'exposició a l'ambient urbà.

Una altra explicació és deguda al fet que la família i l'entorn social són importants en el reconeixement, control i prevenció dels símptomes d'asma (Clarck 1990). Així, l'alta prevalença d'asma en les classes socials més desfavorides podria ser

deguda a un pobre control de l'asma en els pacients d'un nivell baix (Duran-Tauleria 1999), produït per no reconèixer o estar menys preocupat pels símptomes (Ernst 1995), que agreuja els símptomes i fa reduir el percentatge de remissions. De fet, un pobre control de les sibilàncies o xiulets al pit potencialment reversibles està associat amb classe social baixa (Connolly 1989). Les explicacions d'aquesta associació inclouen les característiques dels metges de capçalera i dels pacients, i un pobre accés al sistema sanitari (Duran-Tauleria 1999).

En quasi tots els països de l'estudi hi ha accés universal als serveis de salut, i per tant, sembla que les diferències per classe social no haurien d'estar explicades per un accés diferencial als serveis de salut. No obstant, tal i com va fer notar Ernst (1995), l'accés universal als serveis mèdics no necessàriament significa que la utilització dels serveis sigui igual. Kogevinas (1998) no van trobar diferències en la medicació per l'asma ni en la freqüència de visitar un metge entre asmàtics ocupats i desocupats, tot i que entre els desocupats era menys freqüent haver visitat un especialista. Littlejohns (1989) va reportar que l'adquisició de l'etiqueta d'asmàtic era més freqüent en els pacients de classe social I i II, Lang (1997) va trobar que a Filadèlfia la prescripció d'esteroides inhalats era més baixa en les àrees on un major nombre d'habitants no tenien un diploma d'institut, i Rona (2000) va revisar un quants articles que mostraven que el percentatge d'asmàtics que reben antiinflamatoris és relativament petit en les classes socials baixes.

Tal i com Saldmond (1999) suggeria en el seu article, vam repetir l'estudi de manera que asma fos definit només amb la pregunta "S'ha despertat alguna vegada amb un atac de falta d'aire en els últims 12 mesos?", ja que les altres dues preguntes de la nostra definició, que fan referència a haver tinguts atacs d'asma o haver pres medicació per a l'asma, pressuposen accés als serveis mèdics. Quan vam ajustar el model amb la definició d'una sola pregunta l'efecte de la classe social es mantenia igual que abans, però l'efecte de l'educació era més gran. En canvi, si ajustàvem un model definint asma amb les altres dues preguntes, l'efecte de la classe social es mantenia però el de l'educació desapareixia. Això suggereix que hi ha una relació entre l'ús dels serveis mèdics o el reconeixement dels símptomes d'asma i l'educació, de manera que els de baixa educació són menys conscients de ser asmàtics, i aquest fet pot comportar que

subestimem l'efecte de l'educació quan estudiem asma amb la definició de l'ECRHS basada en tres preguntes.

Tot i que en un estudi transversal com el nostre no podem examinar si la baixa classe social precedeix l'aparició d'asma és molt improbable que l'existència d'una condició com l'asma pugui conduir a un baix nivell socioeconòmic, excepte en casos extrems (Saldmon 1999). No obstant, en un estudi prospectiu (Sibbald 1992), els que havien tingut un historial d'asma o sibilàncies als 23 anys d'edat tenien un risc més alt d'atur, i en el grup d'asmàtics hi havia una proporció més petita de persones que assolien les classes socials I, II o III. Tot i això, l'efecte que trobaven era molt petit, com els mateixos autors reconeixien. Pel que fa l'educació, no hi ha cap evidència que l'asma a la infància redueixi l'educació finalment assolida, i pel que fa a l'exposició laboral a contaminants, és bastant improbable que tenir asma faci que els afectats aconseguixin una feina amb alta exposició a contaminants, sinó que és molt més lògic que passi el contrari.

### **8.1.6. Punts forts i punts febles de l'estudi**

Els punts forts d'aquest estudi són:

- Tenim dades d'un estudi multicèntric que ha fet servir el mateix protocol i les mateixes definicions, cosa que implica que podem treure conclusions més robustes i generals. Aquest fet ens ha permès aplicar models multinivell per valorar l'heterogeneïtat entre centres de les relacions i intentar explicar-la fent servir variables a nivell de centre.
- Es van ajustar les relacions per un conjunt important de possibles confusors de la relació entre asma i nivell socioeconòmic.
- A més a més, es va mesurar un factor molt importat com l'atòpia via IgE i no per qüestionari.

Els punts febles de l'estudi són:

- Possible mala classificació valorant asma per qüestionari. No obstant, les preguntes que s'han fet servir han estat ben validades contra un diagnòstic mèdic d'asma, i de fet, s'ha observat que tenen més validesa que altres mesures suposadament més objectives com un test d'hiperreactivitat bronquial (Burney 1989).
- La mesura de classe social pot tenir alguns problemes, i de fet s'han proposat noves classificacions (Rose 1997) que combinen la professió i l'estatus de treballador, és a dir, si és un treballador amb assalariats, un autònom, un directiu, un supervisor o un empleat. Amb la nova classificació es podrien obtenir resultats una mica diferents, però sembla improbable que les diferències que trobem, que són bàsicament en les categories de treballadors no qualificats (classes IV-V), variïn, ja que en aquestes categories és menys probable que es vegin afectades per l'estatus de treballador.
- Les associacions a nivell de centre estan basades en només 32 centres, i cal ser prudent a l'hora d'interpretar-les. A més a més, el fet de no trobar heterogeneïtat en les relacions entre centres podria ser degut a tenir massa pocs centres (Diez-Roux 2000), tot i que tenim 32 centres mentre que alguns autors han posat el límit inferior del nombre de grups que cal tenir en 30 (Kreft 1996).
- Les nostres dades no diferencien àrees més petites dins d'una mateixa ciutat, cosa que permetria una millor caracterització de l'entorn social de cada individu.
- No resposta: no hi ha relació entre els valors mancants i l'asma quan es fan servir totes les observacions possibles (o sigui, fent servir tots els individus que no tenen missing a la variable "síntomes d'asma en l'últim any", independentment de que tinguin missing en altres variables), cosa que implica l'ús del 98% de les dades. Quan es fa el mateix per classe social i

educació, els de baixa classe social (classes IV-V) i els de més alta educació són els que menys probabilitats de ser inclosos tenen degut a missings a altres variables. No obstant, les relacions crues entre asma i baixa classe social o alta educació són pràcticament iguals quan es fan servir totes les observacions possibles (97%), o quan es fan servir només els individus que finalment han estat inclosos a l'estudi. Finalment, ni el GNP per càpita ni el percentatge d'alta educació en els centres estan relacionats amb el percentatge de no-resposta del centre. Per tant, és molt improbable que la no-resposta hagi afectat significativament els resultats.

## **8.2. Conclusió final**

Els indicadors socioeconòmics, probablement assenyalant exposicions presents i passades a factors ambientals i de comportament, estan associats amb la prevalença d'asma d'una forma homogènia en àrees de països desenvolupats. Aquesta associació no la trobem només a nivell d'individu sinó que també la trobem a nivell de centre, on els individus que viuen en àrees amb més educació tenen un menor risc d'asma, independentment del seu nivell individual d'educació o de classe social.

## **8.3. Treball futur**

Cal remarcar que l'estudi s'hauria de completar, i de fet a l'IMIM hi estem treballant actualment a l'IMIM, separant per asma atòpic i no atòpic, tal i com suggereixen Chen (2002) i Court (2002). També cal estudiar per separat asma sever i asma no sever, com han suggerit altres autors (Volmer 2001, Littlejohns 1993), per tal d'acabar d'entendre com es presenten les diferències socioeconòmiques en l'asma.

Aquesta part no s'ha inclòs en el projecte per la càrrega de treball addicional que suposa, i per tal que la presentació del projecte no s'allargués massa en el temps. No

obstant, la metodologia que es fa servir en aquestes altres fases és tota la mateixa que la que s'ha usat en el projecte, i per tant l'anàlisi que es duria a terme seria el mateix però usant diferents definicions.

# Bibliografía referenciada

Abramson M, Kutin JJ, Raven J, Lanigan A, Czarny D, Walters EH. Risk factors for asthma among young adults in Melbourne, Australia. *Respirology*. 1996 Dec;1(4):291-7.

Alshehri MA, Abolfotouh MA, Sadeg A, Al-Najjar YM, Asindi AA, Al Harthi AM, Al-Trabulsi H, Al Fifi S, Al Frayh A. Screening for asthma and associated risk factors among urban school boys in Abha city. *Saudi Med J* 2000; 21(11):1048-53.

Álvarez-Dardet C, Alonso J, Domingo A, Regidor E. La medición de la clase social en ciencias de la salud. Barcelona: SG Editores. Sociedad Española de Epidemiología, 1995

Anderson D, Aitkin M. Variance Component Models with Binary Response: Interviewer Variability. *Journal of the Royal. Statistical Society B* 1985; 47(2):203-210.

Baltagi BH, *Econometric Analysis of Panel Data*. Wiley, New York, 1995.

Bjornsson E, Hjelm E, Janson C, Fridell E, Boman G. Serology of chlamydia in relation to asthma and bronchial hyperresponsiveness. *Scand J Infect Dis* 1996;28(1):63-9.

Bjornsson E, Hjelm E, Janson C, Fridell E, Boman G. Serology of respiratory viruses in relation to asthma and bronchial hyperresponsiveness. *Ups J Med Sci* 1996;101(2):159-68.

Bjornsson E, Plaschke P, Norrman E, Janson C, Lundback B, Rosenhall A *et al*. Symptoms related to asthma and chronic bronchitis in three areas of Sweden. *European Respiratory Journal* 1994;7:2146-53.

Blalock HM, Wilken PH. *Intergroup Processes: A Micro-Macro Perspective*. New York, NY: Free Press; 1979.

Breslow NE, Clayton DG. Approximate Inference in Generalized Linear Mixed Models. *JASA* 1993; 88(421):9-25.

Bozicevic I, Oreskovic S. Risk factors in asthmatic patients in Croatia. *Coll Antropol* 2000; 24(2):325-34.

Burney P, Malmberg E, Chinn S, Jarvis D, Luczynska C, Lai E. The distribution of total and specific serum IgE in the European Community Respiratory Health Survey. *Journal of Allergy & Clinical Immunology* 1997;99:314-22.

Burney PG, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory Health Survey. *Eur Respir J* 1994; 7(5):954-60.

Burney PG, Laitinen LA, Perdrizet S, Huckauf H, Tattersfield AE, Chinn S, Poisson N, Heeren A, Britton JR, Jones T. Validity and repeatability of the IUATLD (1984) Bronchial Symptoms Questionnaire: an international comparison. *Eur Respir J* 1989; 2(10):940-5.

Burr ML, Verrall C, Kaur B. Social deprivation and asthma. *Respir Med*. 1997; 91(10):603-8.

Chen JT, Krieger N, Van Den Eeden SK, Quesenberry CP. Different slopes for different folks: socioeconomic and racial/ethnic disparities in asthma and hay fever among 173,859 U.S. men and women. *Environ Health Perspect* 2002; 110 Suppl 2:211-6.

Chinn S, Jarvis D, Luczynska CM, Lai E, Burney PG. Measuring atopy in a multi-centre epidemiological study. *Eur J Epidemiol*. 1996;12(2):155-62.

Chowgule RV, Shetye VM, Parmar JR, Bhosale AM, Khandagale MR, Phalnitkar SV *et al*. Prevalence of respiratory symptoms, bronchial hyperreactivity, and asthma in a megacity. Results of the European



- community respiratory health survey in Mumbai (Bombay). *American Journal of Respiratory & Critical Care Medicine* 1998;158:547-54.
- Clark NM, Levison MJ, Evans D, Wasilewski Y, Feldman CH, Mellins RB. Communication within low income families and the management of asthma. *Patient Educ Couns* 1990; 15(2):191-210.
- Connolly CK, Chan NS, Prescott RJ. The influence of social factors on the control of asthma. *Postgrad Med J* 1989; 65(763):282-5.
- Court CS, Cook DG, Strachan DP. Comparative epidemiology of atopic and non-atopic wheeze and diagnosed asthma in a national sample of English adults. *Thorax* 2002; 57(11):951-7.
- de Marco R, Locatelli F, Sunyer J, Burney P. Differences in incidence of reported asthma related to age in men and women. A retrospective analysis of the data of the European Respiratory Health Survey. *American Journal of Respiratory & Critical Care Medicine* 2000;162:68-74.
- De Vocht F, Zock JP, Kogevinas M, Kromhout H. Validity of self-reported occupational exposures in relation to asthma. *Medicina del laboro* 2002; 93(5)
- Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998; 88(2):216-22.
- Diez-Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* 2000; 21:171-92.
- Dogan M, Rokkam S. Introduction. In: Dogan M, Rokkam S, eds. *Social Ecology*. Boston, Mass: MIT Press;1969:1-15.
- Duran-Tauleria E, Rona RJ. Geographical and socioeconomic variation in the prevalence of asthma symptoms in English and Scottish children. *Thorax* 1999; 54(6):476-81.
- ECRHS. Variations in the prevalence of respiratory symptoms, self-reported asthma attacks, and use of asthma medication in the European Community Respiratory Health Survey (ECRHS). *European Respiratory Journal* 1996;9:687-95.
- Engel B, Keen A. A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica* 1994; 48:1-22.
- Ernst P, Demissie K, Joseph L, Locher U, Becklake MR. Socioeconomic status and indicators of asthma in children. *Am J Respir Crit Care Med* 1995; 152(2):570-5.
- Erzen D, Carriere KC, Dik N, Mustard C, Roos LL, Manfreda J, Anthonisen NR. Income level and asthma prevalence and care patterns. *Am J Respir Crit Care Med* 1997; 155(3):1060-5.
- Ferrari M, Poli A, Olivieri M, Tardivo S, Biasin C, Balestreri F *et al*. Seroprevalence of Chlamydia pneumoniae antibodies in a young adult population sample living in Verona. European Community Respiratory Health Survey (ECRHS) Verona. *Infection* 2000;28:38-41.
- Gergen PJ, Mullally DI, Evans R 3rd. National survey of prevalence of asthma among children in the United States, 1976 to 1980. *Pediatrics* 1988; 81(1):1-7.
- Gibbons RD, Bock RD. Trend in correlated proportions. *Psychometrika* 1987; 52:113-24.
- Gibbons RD, Hedeker D. Application of random-effects probit regression models. *Journal of Clinical and Consulting Psychology* 1994; 62: 285-296
- Gibbons RD, Hedeker D. Random-effects probit and logistic regression models for three-level data. *Biometrics* 1997; 53: 1527-37.
- Goldstein H. Nonlinear multilevel models, with an application do discrete response data. *Biometrika* 1991; 78: 45-51.

- Goldstein H. Multilevel Statistical Models. 2<sup>nd</sup> edn. London: Edward Arnold, 1995.
- Gourieroux C, Montfort A. Simulation based econometric methods. Oxford: Oxford University Press, 1996.
- Haan M, Kaplan GA, Camacho T. Poverty and health. Prospective evidence from the Alameda County Study. *Am J Epidemiol.* 1987;125(6):989-98.
- Hamman RF, Halil T, Holland WW. Asthma in schoolchildren. Demographic associations and peak expiratory flow rates compared in children with bronchitis. *Br J Prev Soc Med* 1975; 29(4):228-38.
- Hedeker D, Gibbons RD. MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine* 1996; 49:157-176.
- Higgins BG, Britton JR. Geographical and social class effects on asthma mortality in England and Wales. *Respir Med* 1995; 89(5):341-6.
- Hochstim JR, Athanasopoulos DA, Larkins JH. Poverty area under the microscope. *Am J Public Health Nations Health* 1968; 58(10):1815-27.
- Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley cop ,1989.
- Hox JJ. Applied Multilevel Analysis, Amsterdam: TT-Publications, 1995.
- Isoaho R, Puolijoki H, Huhti E, Kivela SL, Tala E. Prevalence of asthma in elderly Finns. *J Clin Epidemiol* 1994; 47(10):1109-18.
- Janson C, Anto J, Burney P, Chinn S, de Marco R, Heinrich J et al. The European Community Respiratory Health Survey: what are the main results so far? European Community Respiratory Health Survey II. *European Respiratory Journal* 2001;18:598-611.
- Janson C, Chinn S, Jarvis D, Burney P. Physician-diagnosed asthma and drug utilization in the European Community Respiratory Health Survey. *European Respiratory Journal* 1997;10:1795-802.
- Jarvis D, Chinn S, Luczynska C, Burney P. The association of family size with atopy and atopic disease. *Clinical & Experimental Allergy* 1997;27:240-5.
- Jarvis D, Lai E, Luczynska C, Chinn S, Burney P. Prevalence of asthma and asthma-like symptoms in young adults living in three east Anglian towns. *British Journal of General Practice* 1994;44:493-7.
- Kreft IGG. Are Multilevel Techniques Necessary? An overview, Including Simulation Studies. California State University, Los Angeles, 1996.
- Kogevinas M, Anto JM, Sunyer J, Tobias A, Kromhout H, Burney P. Occupational asthma in Europe and other industrialised areas: a population-based study. European Community Respiratory Health Survey Study Group. [erratum appears in *Lancet* 1999 Jul 10;354(9173):166]. *Lancet* 1999;353:1750-4.
- Kogevinas M, Anto JM, Tobias A, Alonso J, Soriano J, Almar E, Muniozguren N, Payo F, Pereira A, Sunyer J. Respiratory symptoms, lung function and use of health services among unemployed young adults in Spain. Spanish Group of the European Community Respiratory Health Survey. *Eur Respir J* 1998; 11(6):1363-8.
- Kuk, A.Y.C. (1995) Asymptotically unbiased estimation in generalized linear models with random effects. *J. R. Statist. Soc. B*, 57, 395-407.
- Lang DM, Sherman MS, Polansky M. Guidelines and realities of asthma management. The Philadelphia story. *Arch Intern Med* 1997; 157(11):1193-200.

- Lewis S, Richards D, Bynner J, Butler N, Britton J. Prospective study of risk factors for early and persistent wheezing in childhood. *Eur Respir J* 1995; 8(3):349-56.
- Lewis SA, Weiss ST, Platts-Mills TA, Syring M, Gold DR. Association of specific allergen sensitization with socioeconomic factors and allergic disease in a population of Boston women. *J Allergy Clin Immunol* 2001;107(4):615-22.
- Litonjua AA, Carey VJ, Weiss ST, Gold DR. Race, socioeconomic factors, and area of residence are associated with asthma prevalence. *Pediatr Pulmonol* 1999; 28(6):394-401.
- Littlejohns P, Macdonald LD. The relationship between severe asthma and social class. *Respir Med* 1993; 87(2):139-43.
- Littlejohns P, Ebrahim S, Anderson R. Prevalence and diagnosis of chronic respiratory symptoms in adults. *BMJ* 1989; 298(6687):1556-60.
- Longford NT. Logistic regression with random coefficients. *Computational Statistics and Data Analysis* 1994; 17:1-15.
- McCulloch CE. Maximum likelihood algorithms for generalized linear mixed models. *JASA* 1997; 92(437): 162-170.
- McWhorter WP, Polis MA, Kaslow RA. Occurrence, predictors, and consequences of adult asthma in NHANESI and follow-up survey. *Am Rev Respir Dis* 1989; 139(3):721-4.
- Mealli F, Rampichini C. Estimating Binary Multilevel Models through Indirect Inference. *Computational Statistics and Data Analysis* 1999; 29: 313-324.
- Meijer E, Leeden R, Busing FMTA. Implementing the Bootstrap Multilevel Models. *Multilevel Modelling Newsletter* 1995; 7(2): 7-11.
- Mielck A, Reitmeir P, Wjst M. Severity of childhood asthma by socioeconomic status. *Int J Epidemiol* 1996; 25(2):388-93.
- Miller JE. Predictors of asthma in young children: does reporting source affect our conclusions? *Am J Epidemiol*. 2001; 154(3):245-50.
- Mitchell EA, Stewart AW, Pattermore PK, Asher MI, Harrison AC, Rea HH. Socioeconomic status in childhood asthma. *Int J Epidemiol* 1989; 18(4):888-90.
- Mitchell EA, Stewart AW. The ecological relationship of tobacco smoking to the prevalence of symptoms of asthma and other atopic diseases in children: the International Study of Asthma and Allergies in Childhood (ISAAC). *Eur J Epidemiol* 2001; 17(7):667-73.
- MlwiN version 1.1. Multilevel models project. London: Institute of Education, University of London.
- Montnemery P, Bengtsson P, Elliot A, Lindholm LH, Nyberg P, Lofdahl CG. Prevalence of obstructive lung diseases and respiratory symptoms in relation to living environment and socio-economic group. *Respir Med* 2001; 95(9):744-52.
- Morgan M, Chinn S. ACORN group, social class, and child health. *J Epidemiol Community Health*. 1983; 37(3):196-203.
- Nascimento-Carvalho CM, Rocha H, Benguigui Y. Effects of socioeconomic status on presentation with acute lower respiratory tract disease in children in Salvador, Northeast Brazil. *Pediatr Pulmonol* 2002; 33(4):244-8.
- Neukirch F, Pin I, Knani J, Henry C, Pison C, Liard R *et al*. Prevalence of asthma and asthma-like symptoms in three French cities. *Respiratory Medicine* 1995;89:685-92.

- Nowak D, Heinrich J, Jorres R, Wassmer G, Berger J, Beck E *et al.* Prevalence of respiratory symptoms, bronchial hyperresponsiveness and atopy among adults: west and east Germany. *European Respiratory Journal* 1996;9:2541-52.
- Pallasaho P, Lundback B, Laspa SL, Jonsson E, Kotaniemi J, Sovijarvi AR, Laitinen LA. Increasing prevalence of asthma but not of chronic bronchitis in Finland? Report from the FinEsS-Helsinki Study. *Respir Med* 1999; 93(11):798-809.
- Odhiambo JA, Ng'ang'a LW, Mungai MW, Gicheha CM, Nyamwaya JK, Karimi F, Macklem PT, Becklake MR. Urban-rural differences in questionnaire-derived markers of asthma in Kenyan school children. *Eur Respir J* 1998; 12(5):1105-12.
- Papageorgiou N, Gaga M, Marossis C, Reppas C, Avarlis P, Kyriakou M, Tsipra S, Zeibecoglou K, Tracopoulos G. Prevalence of asthma and asthma-like symptoms in Athens, Greece. *Respir Med*. 1997;91(2):83-8.
- Peat JK, Woolcock AJ, Leeder SR, Blackburn CR. Asthma and bronchitis in Sydney schoolchildren. II. The effect of social factors and smoking on prevalence. *Am J Epidemiol* 1980; 111(6):728-35.
- Peat JK, Tovey E, Toelle BG *et al.* House dust mite allergens: a major risk factor for childhood asthma in Australia. *Am J Respir Crit Care Med* 1996; 153:141-46.
- Peckham C, Butler N. A national study of asthma in childhood. *J Epidemiol Community Health* 1978; 32(2):79-85.
- Perpiñá, M. *Asma: Lo esencial*. Mra España. 1997.
- Plaschke P, Janson C, Norrman E, Bjornsson E, Ellbjär S, Jarvholm B. Association between atopic sensitization and asthma and bronchial hyperresponsiveness in Swedish adults: pets, and not mites, are the most important allergens. *J Allergy Clin Immunol*. 1999 Jul;104(1):58-65.
- Plaschke PP, Janson C, Norrman E, Bjornsson E, Ellbjär S, Jarvholm B. Onset and remission of allergic rhinitis and asthma and the relationship with atopic sensitization and smoking. *Am J Respir Crit Care Med*. 2000; 162(3 Pt 1):920-4.
- Rodriguez G, Goldman N. An assessment of estimation procedures for multilevel models with binary responses. *J Royal Statistical Society, A*, 1995; 158: 73-90.
- Rona RJ, Hughes JM, Chinn S. Association between asthma and family size between 1977 and 1994. *J Epidemiol Community Health* 1999; 53(1):15-9.
- Rona RJ. Asthma and poverty. *Thorax* 2000; 55:239-44.
- Rose D, O'Reily K. *Constructing classes*. Swindon: ESRC/ONS, 1997.
- Salmond C, Crampton P, Hales S, Lewis S, Pearce N. Asthma prevalence and deprivation: a small area analysis. *J Epidemiol Community Health* 1999; 53(8):476-80.
- Samet JM. Epidemiological approaches for the identification of asthma. *Chest* 1987; 91 (suppl.6): 74S-8S.
- Saxena S, Eliahoo J, Majeed A. Socioeconomic and ethnic group differences in self reported health status and use of health services by children and young people in England: cross sectional study. *BMJ* 2002; 325(7363):520.
- Schwartz J, Gold D, Dockery DW, Weiss ST, Speizer FE. Predictors of asthma and persistent wheeze in a national sample of children in the United States. Association with social class, perinatal events, and race. *Am Rev Respir Dis* 1990;142(3):555-62.
- Sibbald B, Anderson HR, McGuigan S. Asthma and employment in young adults.

Thorax 1992; 47(1):19-24.

SIDRIA (Italian Studies on Respiratory Disorders in Childhood and the Environment). Asthma and respiratory symptoms in 6-7 yr old Italian children: gender, latitude, urbanization and socioeconomic factors. *Eur Respir J* 1997; 10(8):1780-6.

Snijders TAB, Bosker, RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London etc.: Sage Publishers, 1999.

Stewart AW, Mitchell EA, Pearce N, Strachan DP, Weilandon SK. The relationship of per capita gross national product to the prevalence of symptoms of asthma and other atopic diseases in children (ISAAC). *Int J Epidemiol* 2001; 30(1):173-9.

Stiratelli R, Laird N, Ware JH. *Random-Effects Models for Serial Observations with Binary Response*. *Biometrics* 1984; 40: 961-71.

Sunyer J, Anto JM, Kogevinas M, Barcelo MA, Soriano JB, Tobias A *et al*. Risk factors for asthma in young adults. Spanish Group of the European Community Respiratory Health Survey. *European Respiratory Journal* 1997;**10**:2490-4.

Verlato G, Calabrese R, De Marco R. Correlation between asthma and climate in the European Community Respiratory Health Survey. *Arch Environ Health*. 2002; 57(1):48-52.

Volmer T. The socio-economics of asthma. *Pulm Pharmacol Ther* 2001;14(1):55-60.

Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. *Am J Epidemiol* 1986; 123(1): 174-84.

Walters S, Phupinyokul M, Ayres J. Hospital admission rates for asthma and respiratory disease in the West Midlands: their relationship to air pollution levels. *Thorax* 1995; 50(9):948-54.

Watson JP, Cowen P, Lewis RA. The relationship between asthma admission rates, routes of admission, and socioeconomic deprivation. *Eur Respir J* 1996; 9(10):2087-93.

Weiss KB, Gergen PJ, Crain EF. Inner-city asthma. The epidemiology of an emerging US public health concern. *Chest* 1992;101(6 Suppl):362S-367S.

Weitzman M, Gortmaker S, Sobol A. Racial, social, and environmental risks for childhood asthma. *Am J Dis Child* 1990; 144(11):1189-94.

Wieringa MH, Vermeire PA, Van Bever HP, Nelen VJ, Weyler JJ. Higher occurrence of asthma-related symptoms in an urban than a suburban area in adults, but not in children. *European Respiratory Journal* 2001;**17**:422-7.

Wieringa MH, Weyler JJ, Van Bastelaer FJ, Nelen VJ, Van Sprundel MP, Vermeire PA. Higher asthma occurrence in an urban than a suburban area: role of house dust mite skin allergy. *Eur Respir J* 1997;10(7):1460-6.

Wieringa MH, Weyler JJ, Nelen VJ, Van Hoof KJ, Van Bastelaer FJ, Van Sprundel MP, Vermeire PA. Prevalence of respiratory symptoms: marked differences within a small geographical area. *Int J Epidemiol* 1998; 27(4):630-5.

Wong GY, Mason WM. The hierarchical Logistic Regression Model for Multilevel Analysis. *JASA* 1985; 80 (391): 513-24.

World Bank. *World Development Report 1978-1997*. Indexed Omnibus CD-ROM Edition. Asia 2000 Ltd, 1997.

Yemaneberhan H, Bekele Z, Venn A, Lewis S, Parry E, Britton J. Prevalence of wheeze and asthma and relation to atopy in urban and rural Ethiopia. *Lancet* 1997; 350(9071):85-90.

Zeger SL, Karim MR. Generalized linear models with random effects; A Gibbs Sampling Approach. *JASA* 1991; 86:79-86.

Zocchetti C. Estimation of prevalence rate ratios from cross-sectional data. *Int J Epidemiol* 1995; 24(5): 1064-5.

Zocchetti C. Relationship between prevalence rate ratios and odds ratios in cross-sectional studies. *Int J Epidemiol* 1997; 26: 220-3.

Zock JP, Jarvis D, Luczynska C, Sunyer J, Burney P. Housing characteristics, reported mold exposure and asthma in the European Community Respiratory Health Survey. *Journal of Allergy & Clinical Immunology* 2002 A.D.;**110**:285-92.

# Annex 1: Articles de l'ECRHS

## liderats pel centre de Barcelona

(Actualitzada el 8 de gener de 2003)

Basagana X, Sunyer J, Zock JP, Kogevinas M, Urrutia I, Maldonado JA *et al.* Incidence of asthma and its determinants among adults in Spain. *American Journal of Respiratory & Critical Care Medicine* 2001;**164**:1133-7.

Castellsague J, Burgos F, Sunyer J, Barbera JA, Roca J. Prediction equations for forced spirometry from European origin populations. Barcelona Collaborative Group on Reference Values for Pulmonary Function Testing and the Spanish Group of the European Community Respiratory Health Survey. *Respiratory Medicine* 1998;**92**:401-7.

Cortes X, Soriano JB, Sanchez-Ramos JL, Azofra J, Almar E, Ramos J. [European study of asthma. Prevalence of atopy in young adults of 5 areas in Spain. Spanish Group of European Asthma Study]. [Spanish]. *Medicina Clinica* 1998;**111**:573-7.

Cortes X, Soriano JB, Sunyer J, Martinez-Moratalla J, Muniozgoren N, Maldonado JA *et al.* [Factors associated with the development of atopy in young adults]. [Spanish]. *Medicina Clinica* 2000;**114**:165-8.

Galobardes B, Sunyer J, Anto JM, Castellsague J, Soriano JB, Tobias A. Effect of the method of administration, mail or telephone, on the validity and reliability of a respiratory health questionnaire. The Spanish Centers of the European Asthma Study. *Journal of Clinical Epidemiology* 1998;**51**:875-81.

Jaén A, Sunyer J, Basagaña X, Chinn S, Zock JP, Antó JM, Burney P. Specific sensitization to common allergens and pulmonary function in the European Community Respiratory Health Survey. *Clin Exp Allergy* 2000; **32**:1713-9

Kogevinas M, Anto JM, Sunyer J, Tobias A, Kromhout H, Burney P. Occupational asthma in Europe and other industrialised areas: a population-based study. European Community Respiratory Health Survey Study Group. [erratum appears in *Lancet* 1999 Jul 10;354(9173):166]. *Lancet* 1999;**353**:1750-4.

Kogevinas M, Anto JM, Tobias A, Alonso J, Soriano J, Almar E *et al.* Respiratory symptoms, lung function and use of health services among unemployed young adults in Spain. Spanish Group of the European Community Respiratory Health Survey. *European Respiratory Journal* 1998;**11**:1363-8.

Kogevinas M, Anto JM, Soriano JB, Tobias A, Burney P. The risk of asthma attributable to occupational exposures. A population-based study in Spain. Spanish Group of the European Asthma Study. [see comments]. *American Journal of Respiratory & Critical Care Medicine* 1996; **154**:137-43.

Martinez-Moratalla J, Almar E, Sunyer J, Ramos J, Pereira A, Payo F *et al.* [European Asthma Study. Identifying and treating young adults with epidemiological criteria for asthma in five areas of Spain. Spanish Group of the European Asthma Study]. [Spanish]. *Archivos de Bronconeumologia* 1999;**35**:223-8.

Roca J, Burgos F, Sunyer J, Saez M, Chinn S, Anto JM *et al.* Reference values for forced spirometry. Group of the European Community Respiratory Health Survey. *European Respiratory Journal* 1998;**11**:1354-62.

Soriano JB, Anto JM. [An analysis of the prevalence of respiratory symptomatology in the general population]. [letter; comment]. [republished in Arch Bronconeumol 1996 Jun-Jul;32(6):318]. [Spanish]. *Archivos de Bronconeumologia* 1995;31:546.

Soriano JB, Tobias A, Kogevinas M, Sunyer J, Saez M, Martinez-Moratalla J *et al.* Atopy and nonspecific bronchial responsiveness. A population-based assessment. Spanish Group of the European Community Respiratory Health Survey. *American Journal of Respiratory & Critical Care Medicine* 1996;154:1636-40.

Soriano JB, Anto JM, Sunyer J, Tobias A, Kogevinas M, Almar E *et al.* Risk of asthma in the general Spanish population attributable to specific immunoresponse. Spanish Group of the European Community Respiratory Health Survey. *International Journal of Epidemiology* 1999;28:728-34.

Sunyer J, Munoz A. Concentrations of methacholine for bronchial responsiveness according to symptoms, smoking and immunoglobulin E in a population-based study in Spain. Spanish Group of the European Asthma Study. *American Journal of Respiratory & Critical Care Medicine* 1996;153:1273-9.

Sunyer J, Anto JM, Tobias A, Burney P. Generational increase of self-reported first attack of asthma in fifteen industrialized countries. European Community Respiratory Health Study (ECRHS). *European Respiratory Journal* 1999;14:885-91.

Sunyer J, Basagana X, Burney P, Anto JM. International assessment of the internal consistency of respiratory symptoms. European Community Respiratory Health Study (ECRHS). *American Journal of Respiratory & Critical Care Medicine* 2000;162:930-5.

Sunyer J, Kogevinas M, Kromhout H, Anto JM, Roca J, Tobias A *et al.* Pulmonary ventilatory defects and occupational exposures in a population-based study in Spain. Spanish Group of the European Community Respiratory Health Survey. *American Journal of Respiratory & Critical Care Medicine* 1998;157:512-7.

Sunyer J, Anto JM, Kogevinas M, Barcelo MA, Soriano JB, Tobias A *et al.* Risk factors for asthma in young adults. Spanish Group of the European Community Respiratory Health Survey. *European Respiratory Journal* 1997;10:2490-4. Sunyer J, Soriano J, Anto JM, Burgos F, Pereira A, Payo F *et al.* Sensitization to individual allergens as risk factors for lower FEV1 in young adults. European Community Respiratory Health Survey. *International Journal of Epidemiology* 2000; 29:125-30.

Sunyer J, Anto JM, Kogevinas M, Soriano JB, Tobias A, Munoz A. Smoking and bronchial responsiveness in nonatopic and atopic young adults. Spanish Group of the European Study of Asthma. *Thorax* 1997;52:235-8.

Sunyer J, Anto JM, Castellsague J, Soriano JB, Roca J. Total serum IgE is associated with asthma independently of specific IgE levels. The Spanish Group of the European Study of Asthma. *European Respiratory Journal* 1996;9:1880-4.

Tobias A, Soriano JB, Chinn S, Anto JM, Sunyer J, Burney P *et al.* Symptoms of asthma, bronchial responsiveness and atopy in immigrants and emigrants in Europe. European Community Respiratory Health Survey. *European Respiratory Journal* 2001;18:459-65.

Zock JP, Kogevinas M, Sunyer J, Jarvis D, Toren K, Anto JM *et al.* Asthma characteristics in cleaning workers, workers in other risk jobs and office workers. *European Respiratory Journal* 2002;20:679-85.

Zock JP, Jarvis D, Luczynska C, Sunyer J, Burney P. Housing characteristics, reported mold exposure and asthma in the European Community Respiratory Health Survey. *Journal of Allergy & Clinical Immunology* 2 A.D.;110:285-92.

Zock JP, Sunyer J, Kogevinas M, Kromhout H, Burney P, Anto JM. Occupation, chronic bronchitis, and lung function in young adults. An international study. *American Journal of Respiratory & Critical Care Medicine* 2001;163:1572-7.