

# Grau en Estadística

**Títol:** Anàlisi espacial conjunta de càncers de l'aparell digestiu als municipis de Tarragona

**Autor:** Fèlix Bosch Bosch

**Director:** Xavier Puig Oriol

**Departament:** Estadística i Investigació Operativa

**Convocatòria:** Juny 2022





## RESUM

Aquest treball presenta una anàlisi espacial conjunta de la incidència de tres tipus de càncer, d'esòfag, d'estómac i de còlon, en les municipalitats de Tarragona durant el període 2000-2014, distingint per sexes. Aquest estudi el realitzarem amb models bayesians jeràrquics que permeten modelar el comportament espacial i proporcionen evidències de l'existència de diferències en el risc de patir càncer al llarg del territori. Primer estudiarem la incidència dels tres tipus de càncer de forma separada i després els analitzarem de manera conjunta. Els resultats mostren un patró territorial comú entre els càncers d'estómac i d'esòfag, coincidint amb l'anàlisi individual, tant per als homes com per les dones, distingint-lo de la incidència del càncer de còlon. L'anàlisi conjunta ho corrobora en el cas de les dones amb una major dependència de l'efecte espacial compartit, mostrant un efecte espacial igualment distribuït en les tres malalties per al cas dels homes.

## ABSTRACT

This research presents a joint spatial analysis of the incidence of three types of cancer, esophagus, stomach, and colon, in the municipalities of Tarragona during 2000–2014, with gender separation. The study is carried out using hierarchical Bayesian models to approach spatial behavior and disease risk throughout the territory. We will first study the incidence of the three types of cancer individually. Results show a common spatial pattern for stomach and esophagus cancers, in agreement with the individual case, for both men and women, distinguishing it from colon cancer. The joint analysis corroborates it in the case of women with a greater dependency of the shared spatial effect, showing a similar distribution of this effect in the three diseases for men.

Classificació AMS: 62P10.

## INDEX

1. Introducció	2
1.1 El càncer i els estudis estadístics	2
1.2. Els càncers de l'aparell digestiu	4
1.3 Objectius del treball	5
2. Descripció de les dades	6
2.1. Incidència del càncer a Tarragona	6
2.2 Resultats del Mapa del Càncer de Tarragona (2000-2014)	9
3. Models bayesians espacials	11
3.1 Anàlisi espacial bayesià	11
3.2 El Model BYM	12
3.3 Descripció de les dades	13
4. Presentació dels resultats de l'anàlisi per malalties	15
4.1 Resultats sobre el càncer de còlon	15
4.2 Resultats sobre el càncer d'estómac	16
4.3 Resultats sobre el càncer d'esòfag	17
5. Anàlisi espacial conjunt de múltiples malalties	18
5.1 Descripció del model	18
5.2. Resultats del model espacial conjunt	19
6. Conclusions	23
7. Bibliografia	24
Annex. Codis d'R-INLA	25
A1. Codi del model BYM per al cas individual còlon i dones	25
A2. Codi del model BYM per al cas conjunt dones	27

# 1. Introducció

## 1.1 El càncer i els estudis estadístics

Un dels principals problemes de salut en la nostra societat és el càncer. L'Agència Internacional d'Investigació del Càncer, organisme depenent de l'Organització Mundial de la Salut, ha estimat que l'any 2020 s'haurien diagnosticat més de 19.000.000 casos de nous càncers en tot el món i que haurien mort a causa del càncer gairebé 10.000.000 de persones. S'estima que pel 2040 passarem de 19.000.000 de diagnòstics a 30.200.000.

Segons l'Organització Mundial de la Salut (OMS), el càncer és la principal causa de mort en el món: el 2020 aquesta malaltia va causar gairebé 10 milions de defuncions, és a dir, gairebé una de cada sis morts.<sup>1</sup> Els tipus de càncer més comuns són els de mama, pulmó, còlon i recte i pròstata. La mateixa OMS indica que al voltant d'un terç de les morts per càncer es deuen al consum de tabac, a un elevat índex de massa corporal, al consum d'alcohol, a una baixa ingesta de fruites i verdures i a la falta d'activitat física. Aquestes causes estan molt relacionades amb els hàbits socials i culturals, el que justifica els estudis que relacionen la seva incidència amb patrons territorials. L'OMS també afirma que molts casos es poden curar si es detecten a temps i es tracten eficaçment. És, per tant, important conèixer possibles factors relacionats amb la incidència, que podrien permetre gestionar millors les polítiques sanitàries de prevenció.

La incidència d'una malaltia és el nombre de casos d'aquesta que es diagnostiquen durant un període específic de temps en una població determinada, mentre que la mortalitat és el nombre directe de morts. Ambdues mesures s'utilitzen habitualment per estudiar tendències temporals i determinar patrons geogràfics del grau d'exposició i de letalitat dels diferents tipus de càncer.

L'Agència Internacional de Recerca sobre el Càncer indica que els continents amb les taxes d'incidència de càncer més elevades són Europa, Nord-amèrica i Oceania.<sup>2</sup> A Europa, li correspondrien aquest mateix any 2020 gairebé 4.400.000 nous casos de càncer i 1.950.000 defuncions provocades per aquesta malaltia. Aquests valors es van calcular abans d'aparèixer la pandèmia de la Covid-19. El creixement del càncer es deu a diversos factors. Els éssers humans cada cop viuen més anys i a més edat incrementa la probabilitat de desenvolupar alguna forma de la malaltia. També hi ha factors socioeconòmics que entren en joc en molts països de tot el món. Es preveu que el 20% de les persones desenvoluparan càncer en algun moment de la seva vida; un de cada vuit homes moriran per la malaltia i una de cada 11 dones. D'aquí que la prevenció sigui fonamental i motiví els estudis estadístics relacionats amb la incidència.

Segons consta al *Pla Director d'Oncologia de Catalunya* del maig 2018<sup>3</sup>, a Catalunya l'any 2021 es van diagnosticar 32.967 nous casos de càncer, un 2,6% menys respecte el 2019, abans d'aparèixer la pandèmia de la Covid-19. El 2017 va haver uns 23.067 casos incidents de càncer en els homes i 16.095 en les dones. En els homes els més freqüents són els de

---

<sup>1</sup> <https://www.who.int/es/news-room/fact-sheets/detail/cancer>

<sup>2</sup> <https://www.miskawaanhealth.com/cancer/global-cancer-statistics/>

<sup>3</sup> <https://www.idescat.cat/pub/?id=rqp>

pròstata, pulmó, còlon i recte, bufeta de l'orina i els de la cavitat oral i faringe. En el cas de les dones, els més freqüents són els de mama, còlon i recte, coll uterí, pulmó i els limfomes no Hodgkin.

Els estudis epidemiològics són una peça de valor dins la prevenció de la malaltia. Permeten estudiar la incidència de la malaltia segons diversos factors tant de la població (sexe, edat, nivell socioeconòmic, residència, etc.) com del tipus de tumor. Els mapes territorials apareixen com a eines per interpretar patrons geogràfics. Les diferències territorials en la distribució de la incidència poden ser degudes a l'exposició de la població a factors de risc, com l'estil de vida, condicions econòmiques, fenòmens migratoris, tipus de dieta, o directament relacionats amb característiques genètiques.

Els models jeràrquics bayesians són un enfocament molt utilitzat per analitzar les dades espacials de salut pública. Aquestes dades provenen del recompte de casos de malalties en diferents nivells administratius i períodes de temps. Els efectes espacials sovint es modelen fent servir una especificació autoregressiva i condicional (CAR) (Besag, York i Molliee, 1991). Els objectius d'aquests estudis són trobar malalties amb patrons espacials i temporals similars perquè aquests patrons sovint són causats per factors de risc similars. Per tant, en identificar malalties amb patrons similars, es poden descobrir alguns factors de risc compartits. En particular, el consum de tabac i d'alcohol o el tipus de dieta poden tenir condicionants territorials marcats.

L'anàlisi espacial de diverses malalties sovint es basa en models multivariants amb efectes espacials compartits per capturar patrons similars. Per exemple, Knorr-Held i Best (2001) fan servir aquest enfocament per modelar dues malalties considerant un terme espacial compartit en el model amb un pes diferent per a cada malaltia. Downing et al. (2008) proposen un model amb diversos efectes espacials per modelar sis càncers conjuntament. Botella-Rocamora et al. (2015) i Martínez-Beneito et al. (2016) suggereixen un enfocament general per a l'assignació de malalties multivariades que pot ajudar a identificar malalties amb distribucions espacials similars. Marí-Dell'Olmo et al. (2014) utilitzen una anàlisi suavitzada de la variància per a l'anàlisi de diverses malalties en models ecològics.

Pel que fa a estudis més pròxims tenim en particular les recerques de Gómez-Rubio et al. (2019) i Aragonés et al. (2017). Els primers proposen un model bayesià temporal i espacial i l'apliquen al territori espanyol, fent un estudi sobre el càncer d'estómac, esòfag i de cavitat oral. Identifiquen un patró comú espacial entre la cavitat oral i el càncer d'esòfag, i un patró espacial diferent pel càncer d'estómac. També han pogut veure que tots aquests tres tipus de càncer tenen una variació temporal molt similar. L'estudi d'Aragonés et al. (2017) proposa una anàlisi espacial del càncer d'esòfag a Espanya. Fa un estudi per províncies segons el sexe durant els anys 90. Entre homes, l'excés de risc va aparèixer al nord-est i al sud del país. Entre les dones hi havia una notable absència de risc cap al nord i, a diferència dels homes, augmenta en mortalitat, a l'oest d'Extremadura i al sud-est d'Andalusia.

## 1.2. Els càncers de l'aparell digestiu

El càncer de còlon<sup>4</sup> és la tercera forma més comuna de càncer i la tercera causa de mort relacionada amb el càncer en el món occidental. El càncer de còlon causa aproximadament mig milió de morts a l'any. Segons l'Associació Espanyola Contra el Càncer (AECC) i el seu informe *Impacto del càncer en Catalunya 2020*,<sup>5</sup> el 2020, el càncer colorectal es va ser el va provocar més casos d'incidència a Catalunya amb 6135, també va ser el segon tipus de tumor que va provocar més morts, amb 2423, després del de pulmó. El govern de Catalunya està duent a terme un pla de detecció precoç sistemàtic per a la població major de 50 anys des del 2015.

El càncer d'estómac<sup>6</sup> és el cinquè càncer més freqüent en el món, tot i que les taxes globals han anat disminuint. A Catalunya, segons l'informe de l'AECC, és el setè en mortalitat amb 757 al 2020. La seva incidència és molt variable en tots els països, és a dir, varia segons el territori. Per exemple als Estats Units es presenta una incidència de 4,1, a Anglaterra de 3,9 i al Japó de 27,5 per 100 000. Tot i que hi hagi diferències territorials, les variacions respecte al sexe es mantenen en els diferents països, seguint una proporció de dues vegades més freqüents en els homes que en les dones. La incidència més gran per edat es troba entre els 50 i 70 anys, amb una incidència màxima al voltant dels 60 anys, sent infreqüent abans dels 30 anys.

El càncer d'esòfag<sup>7</sup> ocupa el novè lloc, per ordre de freqüència, entre tumors malignes del món, tot i que ocupa el cinquè lloc en els països menys desenvolupats. No apareix en la llista dels tumors més freqüents a Catalunya en l'informe de l'AECC. La seva incidència s'ha multiplicat per 6 en els últims 30 anys amb un rang de supervivència al cap de cinc anys del 15%. Una característica epidemiològica d'aquest tumor, com també passa amb el càncer d'estómac, és la seva gran variabilitat geogràfica. La seva incidència a Europa occidental és de 5 casos per cada 100.000 habitants, mentre que en el sud-est d'Àfrica i en altres àrees com l'Iran o la Xina les taxes d'incidència superen els 100 casos per cada 100.000 habitants. Com també passa amb el càncer d'estómac, predomina en homes, per cada quatre homes que el pateixen, el pateix una dona. Alguns d'aquests tres tipus de càncer tenen factors de risc en comú, com ara el consum d'alcohol, de tabac, el sedentarisme i la dieta nutricional.

---

<sup>4</sup> Càncer colonorectal. (2022, 1 de gener). Viquipèdia, l'Enciclopèdia Lliure. Data de consulta: 1/1/2022 [//ca.wikipedia.org/w/index.php?title=C%C3%A0ncer\\_colorectal&oldid=29058738](https://ca.wikipedia.org/w/index.php?title=C%C3%A0ncer_colorectal&oldid=29058738).

<sup>5</sup> [https://www.contraelcancer.es/sites/default/files/2021-03/Informe%20Impacto%20C%C3%A1ncer%20Catalu%3%B1a\\_2.pdf](https://www.contraelcancer.es/sites/default/files/2021-03/Informe%20Impacto%20C%C3%A1ncer%20Catalu%3%B1a_2.pdf)

<sup>6</sup> Càncer d'estómac. (2022, 1 de gener). Viquipèdia, l'Enciclopèdia Lliure. Data de consulta: 1/1/2022 [//ca.wikipedia.org/w/index.php?title=C%C3%A0ncer\\_d%27est%C3%B3mac&oldid=29058741](https://ca.wikipedia.org/w/index.php?title=C%C3%A0ncer_d%27est%C3%B3mac&oldid=29058741)

<sup>7</sup> Càncer d'esòfag. (2022, 29 de maig). Viquipèdia, l'Enciclopèdia Lliure. Data de consulta: 29/5/2022 [//ca.wikipedia.org/w/index.php?title=C%C3%A0ncer\\_d%27es%C3%B2fag&oldid=30237743](https://ca.wikipedia.org/w/index.php?title=C%C3%A0ncer_d%27es%C3%B2fag&oldid=30237743).

### 1.3 Objectius del treball

L'objectiu d'aquest treball és utilitzar models bayesians espacials per estudiar el comportament de la taxa d'incidència dels tres tipus de càncers de còlon, estómac i esòfag en els municipis de la província de Tarragona durant el període 2000-2014. Plantegem tres objectius específics:

- (1) Observar els patrons individuals de la incidència de cadascun dels càncers distingint per sexes, i veure si es relacionen entre ells, utilitzant un model espacial disjunt per a cada càncer que té en compte, per a cada municipi, tant la incidència pròpia com la del seu veïnatge.
- (2) Utilitzar un model espacial conjunt per veure si apareixen patrons diferents o relacions entre els tres càncers que l'anàlisi disjunta no permetia identificar. Els pesos d'aquest segon model ens ajudaran a interpretar la relació entre els càncers.
- (3) Amb la informació aconseguida i amb el coneixement del comportament espacial dels diferents càncers, avançar algunes hipòtesis sobre la seva distribució territorial i la presència de possibles patrons comuns per a cada sexe.

En les següents seccions d'aquest treball, explicarem l'informe de Galceran et al. (2021) "El mapa de càncer a Tarragona 2000-2014" que està a l'origen de la nostra recerca, tant per la metodologia com per les dades utilitzades. Li segueix un primer estudi basat en un model espacial bayesià que aplicarem de forma individual a les incidències dels tres tipus de càncer. El segon estudi es basa en l'aplicació d'un model espacial bayesià conjunt entre les incidències dels tres tipus de càncer que té en compte els components compartits i específics de cada malaltia. Finalment, compararem els resultats amb els del primer model per veure quina nova informació aportar el model conjunt respecte dels individuals, ampliant així l'estudi sobre el mapa de càncer a Tarragona que prenem com a referència.



## 2. Descripció de les dades

### 2.1. Incidència del càncer a Tarragona

L'informe "El mapa de càncer a Tarragona 2000-2014" (Galceran et al., 2021) ha estat elaborat pel Servei d'Epidemiologia i Prevenció del Càncer de l'Hospital Universitari Sant Joan de Reus. En aquest informe es presenta un estudi de la distribució geogràfica de la incidència i mortalitat dels diferents tipus de càncer en els diferents municipis de la província de Tarragona, distingint per sexe. D'aquesta manera es poden identificar àrees amb risc significativament elevat o baix d'incidència i/o de mortalitat de càncer, agrupant, mitjançant tècniques de clústers geogràfics, zones amb municipis amb incidències o mortalitats significativament elevades o baixes.

Comencem per fer una ullada a la província. En la Figura 1 i taula 1 s'indiquen els municipis de Tarragona i la seva localització.

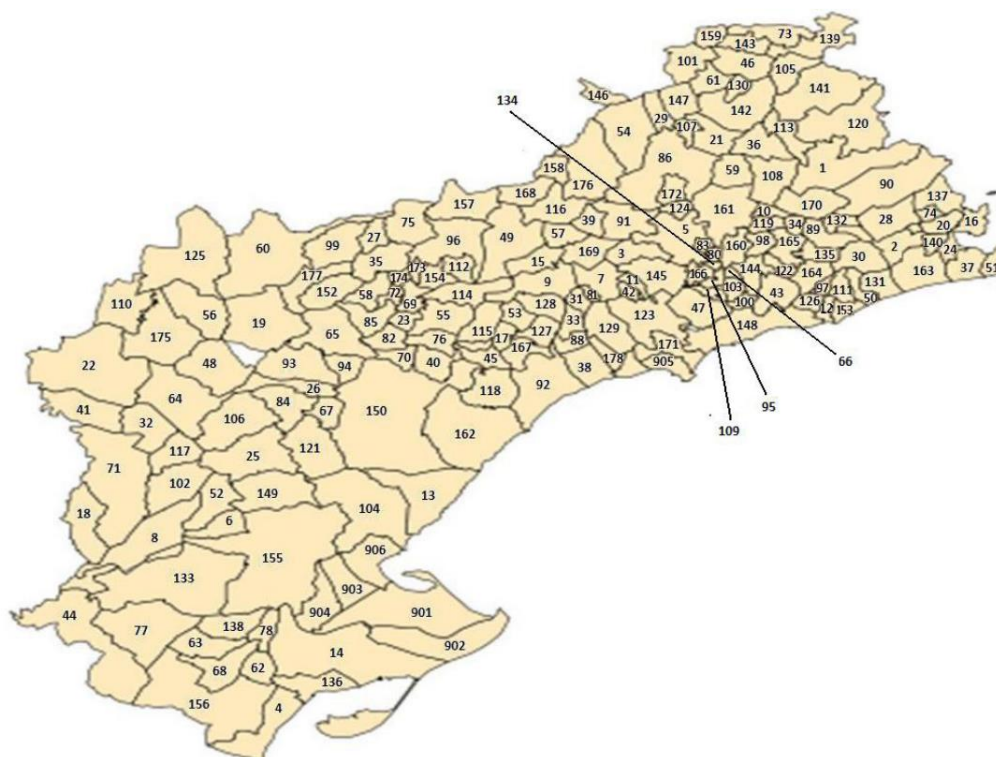


Figura 1. Mapa de Tarragona: comarques i municipis Font: Galceran et al. (2021, p. 10)

Taula 1. Municipis de la demarcació de Tarragona Font: Galceran et al. (2021, p. 11)

1 Aiguamúrcia	2 Albiniana	3 L'Albiol
4 Alcanar	5 Alcover	6 Aldover
7 L'Aleixar	8 Alfara de Carles	9 Alforja
10 Alió	11 Almoster	12 Altafulla
13 L'Ametlla de Mar	14 Amposta	15 Arbolí
16 L'Arboç	17 L'Argentera	18 Arnes
19 Ascó	20 Banyeres del Penedès	21 Barberà de la Conca
22 Batea	23 Bellmunt del Priorat	24 Bellvei
25 Benifallet	26 Benissanet	27 La Bisbal de Falset
28 La Bisbal del Penedès	29 Blancafort	30 Bonastre
31 Les Borges del Camp	32 Bot	33 Botarell

34 Bràfim	35 Cabacés	36 Cabra del Camp
37 Calafell	38 Cambrils	39 Capafonts
40 Capçanes	41 Caseres	42 Castellvell del Camp
43 El Catllar	44 La Sénia	45 Colldejou
46 Conesa	47 Constantí	48 Corbera d'Ebre
49 Cornudella de Montsant	50 Creixell	51 Cunit
52 Xerta	53 Duesaigües	54 L'Espluga de Francolí
55 Falset	56 La Fatarella	57 La Febró
58 Figuera	59 Figuerola del Camp	60 Flix
61 Forès	62 Freginals	63 La Galera
64 Gandesa	65 Garcia	66 Els Garidells
67 Ginestar	68 Godall	69 Gratallops
70 Els Guiamets	71 Horta de Sant Joan	72 El Lloar
73 Llorac	74 Llorenç del Penedès	75 Margalef
76 Marçà	77 Mas de Barberans	78 Masdenverge
79 Masllorenç	80 La Masó	81 Maspujols
82 El Masroig	83 El Milà	84 Miravet
85 El Molar	86 Montblanc	88 Montbrió del Camp
89 Montferri	90 El Montmell	91 Mont-ral
93 Móra d'Ebre	94 Móra la Nova	95 El Morell
96 La Morera de Montsant	97 La Nou de Gaià	98 Nulles
99 La Palma d'Ebre	100 Els Pallaresos	101 Passanant
102 Paüls	103 Perafort	104 El Perelló
105 Les Piles	106 El Pinell de Brai	107 Pira
108 El Pla de Santa Maria	109 La Pobla de Mafumet	110 La Pobla de Massaluca
111 La Pobla de Montornès	112 Poboleda	113 El Pont d'Armentera
114 Porrera	115 Pradell de la Teixeta	116 Prades
117 Prat de Comte	118 Pratdip	119 Puigpelat
120 Querol	121 Rasquera	122 Renau
123 Reus	124 La Riba	125 Riba-roja d'Ebre
126 La Riera de Gaià	127 Riudecanyes	128 Riudecols
129 Riudoms	130 Rocafort de Queralt	131 Roda de Barà
132 Rodonyà	133 Roquetes	134 El Rourell
135 Salomó	136 Sant Carles de la Ràpita	137 Sant Jaume dels Domenys
138 Santa Bàrbara	139 Santa Coloma de Queralt	140 Santa Oliva
141 Pontils	142 Sarral	143 Savallà del Comtat
144 La Secuita	145 La Selva del Camp	146 Senan
147 Solivella	148 Tarragona	149 Tivenys
150 Tivissa	151 La Torre de Fontaubella	152 La Torre de l'Espanyol
153 Torredembarra	154 Torroja del Priorat	155 Tortosa
156 Ulldecona	157 Ulldemolins	158 Vallclara
159 Vallfogona de Riucorb	160 Vallmoll	161 Valls
162 Vandellòs-Hospitalet de l'Infant	163 El Vendrell	164 Vespella de Gaià
165 Vilabella	166 Vilallonga del Camp	167 Vilanova d'Escornalbou
168 Vilanova de Prades	169 Vilaplana	170 Vila-rodona
171 Vila-seca	172 Vilaverd	173 La Vilella Alta
174 La Vilella Baixa	175 Vilalba dels Arcs	176 Vimbodí
177 Vinebre	178 Vinyols i els Arcs	901 Deltebre
902 Sant Jaume d'Enveja	903 Camarles	904 L'Aldea
906 L'Ampolla	92 Mont-roig del Camp	905 Salou

La Figura 2 mostra un mapa de la província de Tarragona amb les seves 10 comarques: Alt Camp, Baix Camp, Baix Ebre, Baix Penedès, Conca de Barberà, Ribera de l'Ebre, Montsià, Priorat, Tarragonès i Terra Alta. Comprèn un total de 814.300 habitants distribuïts de forma molt desigual en 183 municipis. Els més importants són Tarragona i Reus, que reuneixen 240.000 habitants, el que vindria a ser aproximadament un 29% de la població i estan situats a la zona de la costa, a 12 km l'un de l'altre. Els municipis més extensos són Tortosa i Tivissa, superant els 200 kilòmetres quadrats cadascun d'ells. En les comarques de Tarragona i Baix Camp és on se situa la vida més industrial i de ciutat. Les comarques interiors, Priorat, Ribera d'Ebre, Conca de Barberà, etc., hi trobem una vida més rural i de camp.



Figura 2. Mapa de les comarques de Tarragona<sup>8</sup>

La província de Tarragona té dues centrals nuclears: una situada al municipi d'Ascó i l'altra al municipi de Vandellós, com veiem en la Figura 3. Avui en dia està demostrat l'efecte que tenen les centrals nuclears en les persones. En el cas del càncer, provoquen sobretot càncer de tiroides i també poden provocar tumors cerebrals i d'ossos. En principi, no haurien d'afectar l'estudi de càncer de l'aparell digestiu, però per si de cas o tindrem en compte, ja que potser ens dona resposta a certs patrons o punts extrems que observem en l'estudi.

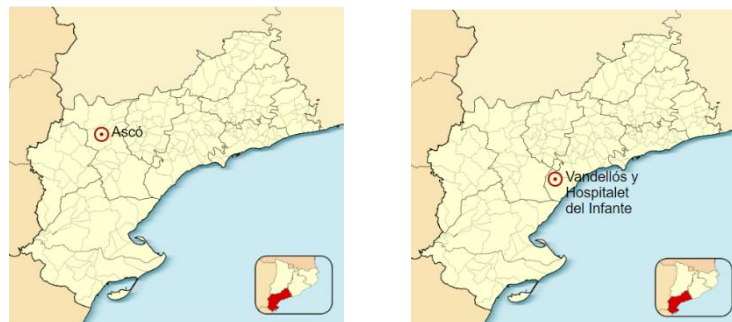


Figura 3. Centrals nuclears a la província de Tarragona<sup>9,10</sup>

La província de Tarragona compta amb dues regions sanitàries: Terres de l'Ebre (amb les comarques de Baix Ebre, Montsià, Ribera d'Ebre i Terra Alta) i Camp de Tarragona (amb les comarques d'Alt Camp, Baix Camp, Baix Penedès, Conca de Barberà, Priorat i Tarragonès). La informació sobre els nous casos d'incidència o de mortalitat de càncer s'obté dels centres sanitaris d'aquestes dues regions sanitàries, siguin públics o privats.

L'any 1979 es va crear el *Registre de Càncer de Tarragona* (RCT) que abastava tota la demarcació provincial de Tarragona i que va començar a proporcionar dades d'incidència de càncer de manera regular. Això va possibilitar, per primera vegada a Catalunya, el coneixement de la magnitud del càncer a la nostra població. El RCT és un registre de càncer de base poblacional amb l'objectiu d'identificar i enregistrar tots els nous casos d'incidència i

8 <https://www.diarimes.com/pags/comarques.html>

9 <https://en.wikipedia.org/wiki/Asc%C3%B3>

10 [https://es.wikipedia.org/wiki/Vandell%C3%B3s\\_y\\_Hospitalet\\_del\\_Infante](https://es.wikipedia.org/wiki/Vandell%C3%B3s_y_Hospitalet_del_Infante)

de mortalitat de càncer entre els habitants residents a la demarcació provincial de Tarragona (regions sanitàries Camp de Tarragona i Terres de l'Ebre), independentment del lloc on s'hagin diagnosticat. És a dir, si algú viu a Reus i se'n va de viatge el Japó i un cop allà se li diagnostica un càncer, serà enregistrat pel RCT.

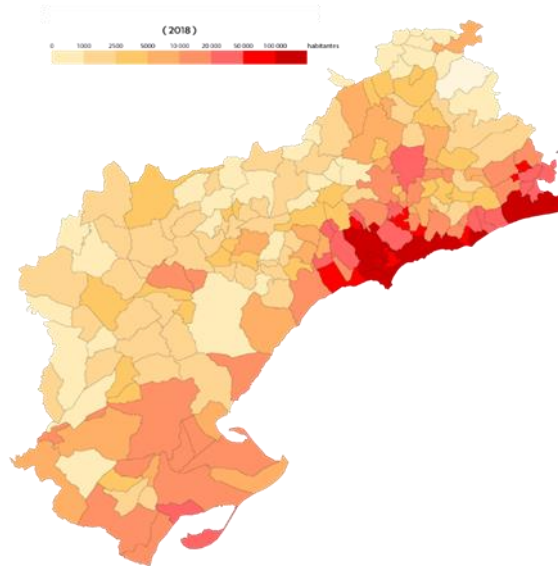
La seva activitat comporta la recollida de dades sistemàtica, continuada i exhaustiva. Aquestes dades ajuden a l'elaboració de resultats per determinar la incidència, la mortalitat, la supervivència i la prevalença del càncer. També ajuden a obtenir distribucions segons múltiples variables, així com l'evolució en el temps i l'estimació futura d'aquests indicadors. Tota la informació sobre el càncer de la població contribueix a la planificació de serveis sanitaris, avaluar les intervencions sanitàries tant a escala preventiva com assistencial i col·laborar en estudis etiològics i analítics.

El Registre de Càncer de Tarragona (RCT) ha estimat que l'any 2021 es diagnosticaran més de 4.700 nous càncers. El càncer és la causa d'un 34% de la mortalitat dels homes i d'un 22% de les dones a les terres tarragonines. És també la causa d'una pèrdua de qualitat de vida i d'un augment de patiment que avui en dia no som capaços de quantificar. El RCT segueix els protocols internacionals que abasten des de la definició de cas fins als seus sistemes d'operació i l'elaboració de resultats. Aquestes normes i criteris s'han desenvolupat per tal de garantir la fiabilitat i la comparabilitat de les dades entre els diferents registres de càncer de base poblacional de tot el món.

El RCT considera com a cas registrable qualsevol tumor maligne invasiu o no. També considera registrables tots els tumors del sistema nerviós central benignes i de comportament no especificat de qualsevol edat. Tot això sempre que les persones diagnosticades siguin residents a la demarcació provincial de Tarragona en el moment del diagnòstic o en el moment de la defunció (si el certificat de defunció és l'única font d'informació).

## 2.2 Resultats del Mapa del Càncer de Tarragona (2000-2014)

Galceran et al. (2021) han trobat evidències de l'existència de diferències significatives en el risc de patir càncer al llarg del territori de Tarragona i que la incidència està relacionada amb la mida de la població. Com més habitants té una població més alta és la taxa d'incidència, i varia segons el sexe. En el cas dels homes tenen una taxa d'incidència d'un 26% més elevada a les ciutats amb més de 50.000 habitants respecte a els municipis amb menys de 2.500 habitants. En el cas de les dones, la diferència és del 15% i està al límit de la significació estadística. Per tant, com observem en la Figura 4, és lògic que els municipis del triangle Tarragona-Reus-Valls i alguns municipis adjacents siguin les zones amb major incidència de càncer. En conseqüència, a la Regió sanitària Camp de Tarragona s'observa una major incidència de càncer que a la Regió sanitària Terres de l'Ebre.



*Figura 4. Densitat de la població de Tarragona<sup>11</sup>*

Aquesta situació es dona sobretot en els casos de tumors associats a una vida més urbana i industrial, ja que aquesta va lligada a una prevalença més elevada de tabaquisme, consum d'alcohol, sobrepès, dieta pobra en vegetals (fruites i verdures) i contaminació, entre altres factors. Aquests casos s'accentuen amb certs tipus de càncer com el d'esòfag, mama (en dones), còlon i recte, pàncrees (en homes), pulmó, laringe, pròstata i bufeta urinària.

L'informe també comenta que alguns tipus de càncer són més freqüents en la Regió sanitària Terres de l'Ebre que al Camp de Tarragona, com el melanoma de la pell i la leucèmia en ambdós sexes, i els càncers de coll uterí, cos uterí, ovari i encèfal en les dones. No sempre es troba un patró per cada tipus de càncer: hi ha casos difícils d'estudiar a causa de la poca incidència o a la poca informació disponible al respecte. D'aquí l'interès que tenen els models espacials conjunts com el que presentem en aquest treball, doncs permeten establir patrons comuns a diferents malalties i, per tant, poder inferir les incidències d'unes a partir de les altres.

<sup>11</sup> [https://es.wikipedia.org/wiki/Provincia\\_de\\_Tarragona](https://es.wikipedia.org/wiki/Provincia_de_Tarragona)

## 3. Models bayesians espacials

### 3.1 Anàlisi espacial bayesià

Tant l'estudi de Galceran et al. (2021) que agafem com a referència, com el que presentem aquí es basen la utilització de models espacials bayesians jeràrquics. Aquests models són l'enfocament més habitual per tractar dades espacials de malalties.

Els models bayesians parteixen del principi que els paràmetres de qualsevol funció de probabilitat tinguin distribucions, és a dir, siguin variables aleatòries. A aquestes distribucions se'n diu distribucions a priori. A més a més, tal i com passa en el model espacial, els paràmetres de les distribucions a priori esmentades poden ser també variables aleatòries, d'aquí que s'estableixi una jerarquia i es parli de model jeràrquic. Combinant la versemblança amb les distribucions a priori dels paràmetres, s'obté la distribució a posteriori que descriu el comportament dels paràmetres considerant la informació que aporten les dades recollides.

Com a avantatges de l'estadística bayesiana podem dir que permet la formulació de models més complexos i flexibles que l'estadística freqüentista i que el programari bayesià ha aconseguit en l'actualitat un grau de maduresa que permet la utilització quotidiana d'aquesta metodologia. Si en l'estadística clàssica només es pren com a font d'informació les mostres finites obtingudes, en el cas bayesià, a més de la mostra també juga un paper fonamental la informació prèvia dels fenòmens que es tracten de modelitzar: les distribucions a priori.

Pel que fa a l'anàlisi espacial, la representació cartogràfica de malalties s'utilitza comunament per descriure la distribució geogràfica dels índexs d'incidència o mortalitat de malalties, generant així hipòtesis sobre les seves possibles causes i diferències. Els models jeràrquics que es fan servir es basen en una regressió de Poisson que inclou efectes espacials utilitzant una especificació autoregressiva condicional (CAR). L'ajustament del model bayesià es duu a terme utilitzant mètodes Monte Carlo de cadenes de Markov (MCMC), i també utilitzant el mètode INLA (Integrated Nested Laplace Approximation) que requereix menys temps computacional i que és el que es farà servir en aquest treball.

El model bayesià espacial més usat és el proposat per Besag, York i Mollie (BYM). Aquest model és la principal referència en mètodes de suavitzat de riscos. El model BYM estableix una ponderació entre dues fonts de variació per a obtenir una estimació del risc relatiu de cada àrea. La primera d'aquestes fonts de variació, d'estructura espacial, comparteix informació entre unitats veïnes per a modelitzar la dependència geogràfica dels riscos. Considerarem que una àrea és veïna d'una altra si comparteix frontera. Aquest efecte fa possible que regions veïnes tinguin estimacions similars i d'aquesta manera el risc varia geogràficament de manera suau. El segon dels efectes, espacialment heterogeni, pren valors independents en totes les unitats geogràfiques, cosa que permet que localitzacions veïnes presentin riscos diferents. En el cas que una àrea petita tingui una població gran, tindrà més pes la informació proporcionada per aquesta àrea; en canvi, si presenta una població de grandària reduïda no tindrà tant pes com la resta d'àrees (o àrees veïnes).

Aquest enfocament reforça la inferència de cada regió usant informació de les regions veïnes i permet identificar, entre les malalties relacionades, subpoblacions amb factors etiològics comuns. Segons Otiende et al. (2020), la informació que prové de les regions veïnes permet observar la concurrència de patrons i el condicionament d'una malaltia en altres. Aquest coneixement pot ser molt valuós en casos de malalties no comunes on hi ha incertesa de recompte i/o falta d'informació.

### 3.2 El Model BYM

Pels models individuals de cada malaltia, hem seguit la proposta de Navarro (2012) que utilitza models BYM amb el mètode INLA (Integrated Nested Laplace Approximation) i el paquet de R associat R-INLA. Veiem primer els principis del model BYM.

Sigui  $\theta_i$  el risc relatiu de la malaltia en cada municipi (en el nostre cas,  $i = 1, \dots, 183$ ) segons el sexe (per als homes o per les dones).  $O_i$  és el nombre de casos observats d'incidència de la malaltia durant el període de l'estudi (2000-2014).

Suposem que  $O_i$  segueix una distribució Poisson amb valor esperat  $E_i\theta_i$ , on  $E_i$  és el nombre de casos esperats a la província  $i$ -èsima suposant risc constant, és a dir, basat només en la taxa d'incidència global i en la població de risc al municipi  $i$ -èsim ajustada per edat i sexe:

$$O_i | \theta_i \sim P(E_i\theta_i)$$

$$\log(\theta_i) = \alpha + u_i + v_i$$

El logaritme del risc relatiu es modelitza de forma lineal on  $\alpha$  és el terme independent. El terme  $u_i$  recull l'efecte espacial: si dues àrees són veïnes tindran valors més "semblants". El terme  $v_i$  són els efectes independents aleatoris i representa la variabilitat no estructurada, és a dir, el "soroll" dels efectes aleatoris no espacials.

Seguint Navarro (2012), suposem que les variables  $u_i$  i  $v_i$  son independents i que el factor soroll  $v_i \sim N(0, \sigma_v^2)$  sent  $\sigma_v^2$  la variància desconeguda. Per a  $u_i$  utilitzarem el que es coneix com *autorregressió intrínseca gaussiana* (CAR):

$$u_i \sim N\left(\sum_{i \sim j} u_j/n_i, \sigma_u^2/n_i\right)$$

on  $i \sim j$  indica que  $i$  i  $j$  són veïns. Aquesta aproximació especifica la dependència espacial com una mitjana de l'efecte de les àrees veïnes on  $n_i$  és el número de veïns de  $i$ . Com a resultat final obtenim per a cada municipi la mitjana a posteriori de la distribució de risc i la probabilitat que sigui superior a 1.

Al terme constant  $\alpha$  li assignem un a priori no informativa (improper flat prior), és a dir, un a priori que no aportarà informació al model. En el nostre serà una distribució uniforme de 0 a 10. La priori de  $\sigma_u^2$  és una LogGamma amb  $a = 1, b = 0.001$ .

En la pràctica, una vegada obtingudes les estimacions a posteriori dels paràmetres, s'estimen els riscos de cada província a través de la següent expressió:

$$RR_i = \theta_i = \exp(\alpha_i + u_i + v_i)$$

El risc relatiu ( $RR_i$ ) compara el risc de cada província amb el nivell mitjà de Tarragona.

### 3.3 Descripció de les dades

Les dades que considerem ens les ha facilitat el mateix Registre de Càncer de Tarragona que duu a terme el Servei d'Epidemiologia i Prevenció del Càncer de l'Hospital Universitari Sant Joan de Reus. Aquestes dades ens proporcionen els casos observats i els casos esperats de cada tipus de malaltia segons el municipi de Tarragona des de l'any 2000 fins al 2014.

A l'hora de calcular els casos esperats s'ha fet de la següent manera:

$$E_i = \sum_j P_{ij} \lambda_j$$

on  $P_{ij}$  és la població del municipi  $i$ , pel grup d'edat  $j$ .  $\lambda_j$  fa referència a la taxa per a cada combinació de tipus tumoral i sexe pel grup d'edat  $j$  en el global de Tarragona. La taxa de referència es calcula agafant els casos observats per a cada combinació de tipus de càncer, edat i sexe i dividint-los entre la població total de la combinació en qüestió.

La nostra base de dades conté set variables:

- Municipi: variable numèrica que cataloga els municipis
- Codi: codi postal del municipi
- Nom: indica el nom del municipi
- Obs: nombre de casos d'incidència del tumor observats
- Esp: nombre de casos d'incidència del tumor esperats
- Causa: el tipus de tumor
- Sexe: home o dona

La Figura 9 presenta les distribucions dels casos d'incidència en homes (esquerra) i dones (dreta) dels tres tipus de càncer. El resum numèric el trobem a les Taula 1 i 2. Observem que el càncer de còlon és el que presenta més casos d'incidència en els dos sexes, d'una manera significativa, sobretot si mirem els valors màxims. En el cas dels homes, el 75% dels municipis estan per sota les 17 incidències, però n'hi ha dos que superen els 750. Pel que fa a les dones, el 75% dels municipis no superen les 12 incidències però n'hi ha dos que superen les 450. Pel que fa al càncer d'estómac i esòfag, el 50% dels municipis tenen una incidència o cap, tant pels homes com per les dones. De l'altra meitat de municipis, on hi ha més casos d'incidència, el càncer d'estómac té valors més alts tant en dones com en homes. De fet, si en fixem en els valors màxims i trobem diferències bastant significatives.



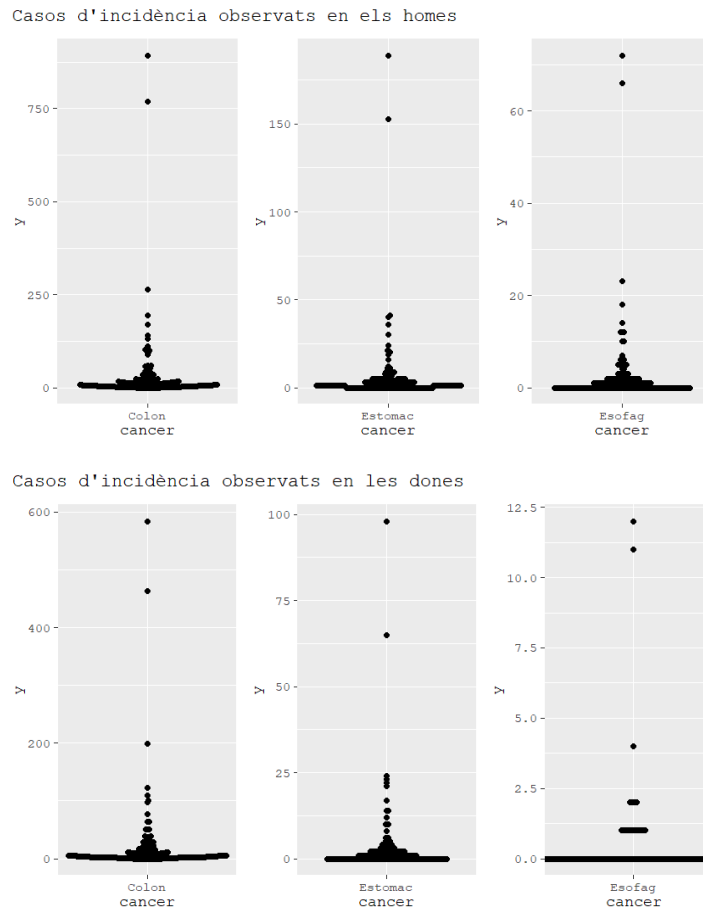


Figura 5 Casos d'incidència observats en homes (a dalt) i dones (a baix).  
Font: elaboració pròpia

Taula 2. Casos d'incidència observats en els homes. Font: elaboració pròpia

Càncer de	Mínim	Primer quantil	Mediana	Tercer quantil	Màxim
Còlon	0	3	7	17	893
Estómac	0	0	1	4	189
Esòfag	0	0	1	2	72

Taula 3. Casos d'incidència observats en les dones Font: elaboració pròpia

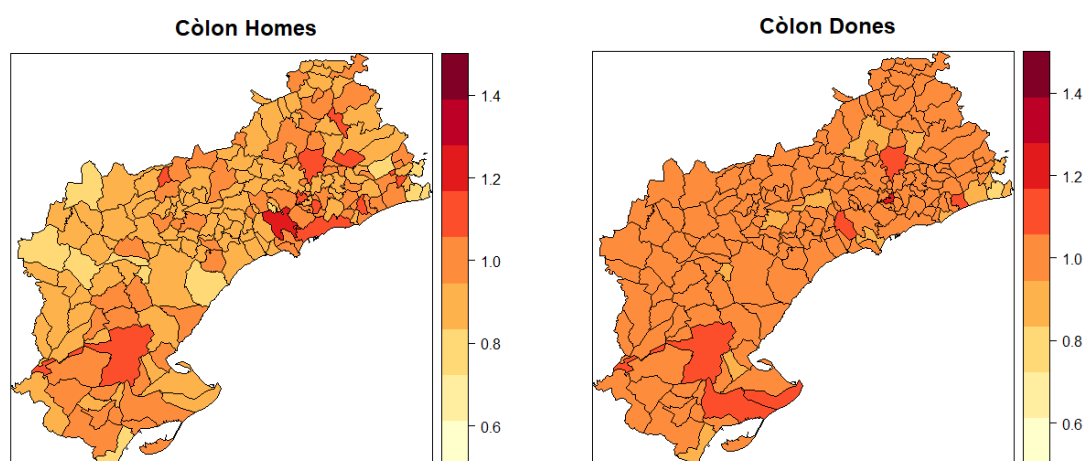
Càncer de	Mínim	Primer quantil	Mediana	Tercer quantil	Màxim
còlon	0	2	4	12	584
Estómac	0	0	1	2	98
Esòfag	0	0	0	0	12

## 4. Presentació dels resultats de l'anàlisi per malalties

Per elaborar el model individual per a cada tipus de càncer distingint per sexes, hem utilitzat la proposta de Navarro et al. (2012) així com el codi que faciliten aquests autors en el seu article (vegeu Annex A1 per al cas del càncer de còlon i les dones).

Un cop aplicat el model BYM de forma individual a cada tipus de càncer, distingint per sexe, obtenim els resultats i hi projectem els valors esperats de l'índex d'incidència estandarditzat, on el valor 1 correspon al valor mitjà de la província de Tarragona: valors inferiors a 1 indiquen poc risc; superiors a 1 indiquen un risc més alt que la mitjana de la província.

### 4.1 Resultats sobre el càncer de còlon

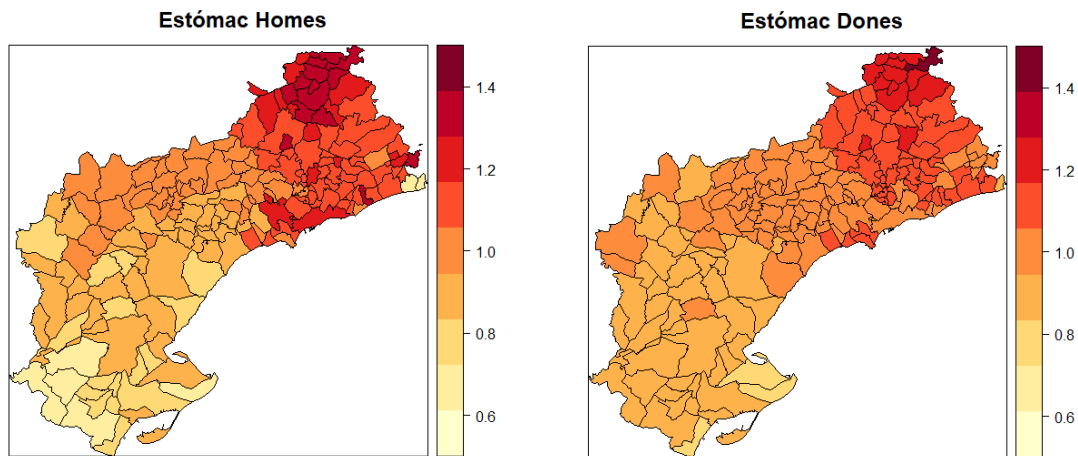


*Figura 6. Casos d'incidència. Font: elaboració pròpia*

Començarem pels resultats del càncer de còlon. Com observem en la Figura 6, en el cas dels homes, no hi ha un clar patró determinat, sinó que en general veiem poc risc excepte punts en concert. En primer lloc, hi trobem zones de risc a la zona més urbana de la província, sobretot en Tarragona i Reus i als seus voltants. També trobem una zona de risc més cap al sud, als municipis de Roquetes i Tortosa.

En les dones es veu un comportament similar, però amb major incidència cap al Delta de l'Ebre en municipis com Tortosa, Deltebre i Amposta, i menor incidència a la zona de Tarragona i Reus. La major incidència en zones més poblades i industrialitzades es pot explicar pels factors relacionats amb l'estil de vida urbà (sedentarisme, tabaquisme, dieta, etc.). En general les dones tenen una taxa d'incidència més alta que els homes, tot i que, en el seu cas no tenen punts tan extrems en el cas de Reus, però, d'altre banda, cap al sud hi té dos municipis amb una taxa més alta, Amposta i Sant Jaume d'Enveja.

## 4.2 Resultats sobre el càncer d'estómac

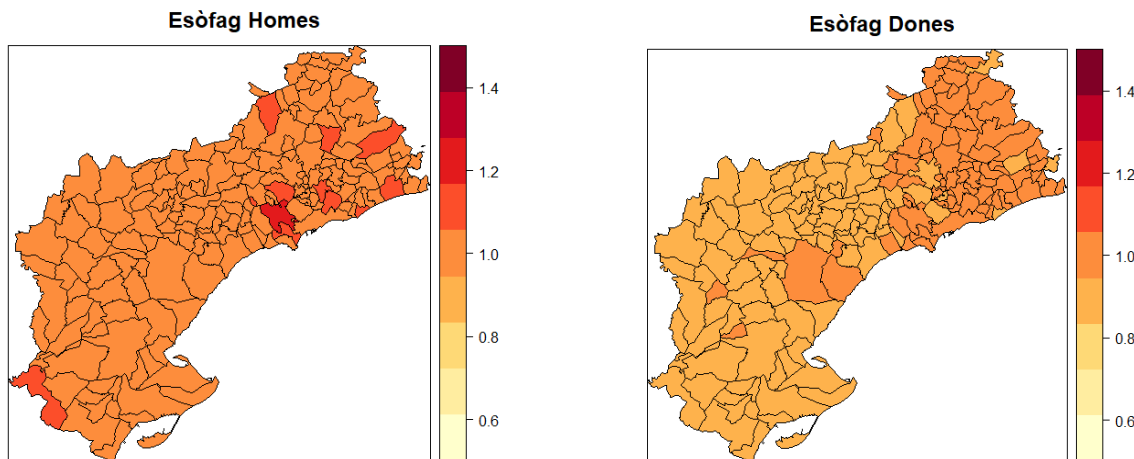


*Figura 7. Casos d'incidència. Font: elaboració pròpia*

En el cas del càncer d'estómac, en la Figura 7 observem un patró claríssim, on tenim la zona del Delta de l'Ebre, el sud-oest, amb una taxa d'incidència per sota d'1, baix risc, i la zona més cap al nord-est amb alt risc, on la majoria de municipis tenen unes taxes d'incidència que superen el 1,2. Aquest patró el comparteixen tant homes com dones. Com a diferència, observem que el patró dels homes és més extrem, és a dir, zones amb risc molt baix i zones amb risc molt alt, d'altra banda, les dones tenen un patró més 'suau'. Les taxes més elevades es presenten a la Conca de Barberà, d'una manera força destacada, seguit de l'Alt Camp, el Tarragonès i el Baix Penedès. Al Baix Ebre i al Montsià és on hi trobem les taxes d'incidència més baixes.

Curiosament, la zona amb les incidències més altes i la zona amb incidències més baixes són zones rurals. Potser seria interessant destinar esforços a intentar esbrinar-ne la causa o causes.

### 4.3 Resultats sobre el càncer d'esòfag



*Figura 8. Casos d'incidència. Font: elaboració pròpia*

Per última tenim el cas del càncer d'esòfag. Com veiem en la Figura 8, a diferència dels altres dos, en el cas del càncer d'esòfag hi ha una taxa d'incidència més alta en homes que en dones, cosa que ja hem comentat anteriorment i ha quedat confirmat aquí. Podríem considerar que en aquest cas el patró és similar al del càncer d'estómac però d'una manera molt menys exagerada.

S'observa que les taxes d'incidència són lleugerament més elevades a la Regió sanitària Camp de Tarragona que a la de les Terres de l'Ebre, especialment en el cas de les dones. De fet, s'observa un patró geogràfic en el que la incidència menys elevada es presenta a la zona sud-oest i va augmentant a mesura que ens traslладem cap a la zona nord-est del territori (Baix Penedès, Tarragonès, Conca de Barberà i part est del Baix Camp). L'única cosa diferencial és la taxa més elevada de la ciutat de Reus en els homes.

En síntesi, pel que fa als tres càncers considerats, podem observar un comportament similar, tant pel que fa als homes com a les dones, dels càncers d'estómac i d'esòfag, amb una incidència molt més gran del primer i un contrast més marcat entre les zones nord i sud del territori. La incidència del càncer de còlon, en canvi, presenta una distribució territorial diferent. Les diferències entre homes i dones són baixes en tots tres casos.

## 5. Anàlisi espacial conjunt de múltiples malalties

### 5.1 Descripció del model

Per fer el model conjunt dels tres tipus de càncers considerats, utilitzem la proposta de model conjunt espacial-temporal de Gómez-Rubio et al. (2019) que nosaltres restringim al cas espacial (sense dimensió temporal) donat que treballem amb dades agregades 2000-2014 i considerant tres malalties. Segons aquests autors, el model BYM es pot aplicar a diferents malalties de forma conjunta per produir estimacions espacials del risc d'incidència i fer anàlisis cartogràfiques que permetin identificar patrons de risc comuns. L'anàlisi conjunta permet obtenir millors estimacions d'aquests patrons comuns i identificar desviacions específiques en algunes àrees.

El fet de considerar informació de malalties relacionades ajuda a quantificar la variabilitat esperada relacionada, per exemple, quan hi ha factors de risc compartits. En la planificació de la salut, trobar un patró comú, per exemple, entre dues causes de mort "evitables" permetria noves actuacions del sistema sanitari. En un esperit similar, l'anàlisi conjunta pot descobrir patrons anòmals derivats d'una font de recollida de dades comuna.

Seguirem aquí la proposta de model conjunt de Gómez-Rubio et al. (2019) per al cas de tres malalties (que es generalitza fàcilment a un nombre qualsevol  $D$ ). En un model conjunt, el risc relatiu inclou termes compartits per les diferents malalties. A l'efecte en qüestió se li poden aplicar diferents pesos que mesurin la distribució geogràfica de la malaltia respecte del patró comú. En aquest model conjunt, tindrem el següent:

$$\begin{aligned} O_i^{(d)} | E_i^{(d)}, \theta_i^{(d)} &\sim Po(E_i^{(d)} \theta_i^{(d)}), d = 1, 2, 3 \\ \log(\theta_i^{(d)}) &= \alpha^{(d)} + \delta^{(d)} U_i + u_i^{(d)} \\ \delta^{(1)} * \delta^{(2)} * \delta^{(3)} &= 1 \\ d = 1 &\rightarrow C\grave{o}lon, d = 2 \rightarrow Est\acute{o}mac, d = 3 \rightarrow Es\grave{o}fag \end{aligned}$$

on  $\alpha^{(d)}$  son els termes independents específics de cada malaltia,  $U_i$  és el patró espacial compartit amb els pesos  $\delta^{(d)}$  corresponents i  $u_i^{(d)}$  son patrons espacials específics de cada malaltia.

Si ho comparem amb el model anterior, l'individual, simplement hem afegit un nou terme  $U_i$  que assigna una influència major o menor al risc segons el pes assignat,  $\delta^{(d)}$ .

Els vectors específics i els efectes compartits es defineixen de la següent manera:

$$u^{(d)} \sim CAR(W, \tau_d) \quad d = 1, 2, 3; \quad U \sim CAR(W, \tau_0)$$

on  $W$  és la matriu dels veïnatges. En aquest model  $U$  està multiplicada pels pesos, això pot causar problemes de confusió entre els pesos i l'escala dels efectes, és a dir, la precisió  $\tau_0$ . Per tant, considerarem que  $\tau_0$  és igual a 1, per evitar problemes.

Pel que fa a les distribucions a priori de la resta de paràmetres, hi ha diferents opcions possibles. Seguint sempre la proposta de Gómez-Rubio et al. (2019), i com hem fet en el model previ, als termes constants  $\alpha^{(d)}$  específics de cada malaltia els assignem un a priori no informativa (*improper flat prior*), és a dir, un a priori que no aportarà informació al model. Als pesos espacials se'ls assigna una log-Normal amb mitjana zero i precisió 1/5.9 (valor que els autors atribueixen a Downing et al. (2008)).

Partim doncs d'uns pesos positius amb quantils 0,25 i 0,975 de 0,0086 i 116,8319 respectivament. Això significa que els pesos poden prendre valors molt petits, indicant un efecte insignificant dels termes espacials compartits en el predictor lineal. Per tant, no és necessari que les malalties del model es correlacionen per endavant, ja que el model pot produir pesos molt petits en aquest cas. És clar que llavors caldria reconsiderar la inclusió de les malalties no correlacionades en el model.

Per als paràmetres d'escala dels efectes aleatoris, Gómez-Rubio et al. (2019) suggereixen provar diferents prioritats per tal de elaborar una anàlisi de sensibilitat sobre els resultats i investigar com diferents priors impacten en les estimacions dels riscos relatius i altres paràmetres en el model. Així, proposen ajustar tres models diferents en els quals tots els paràmetres d'escala tinguin les mateixes distribucions. En primer lloc, suggereixen una distribució uniforme entre 0 i 10 sobre les desviacions estàndard; una semi-Cauchy (amb paràmetre 25); o una gamma invertida amb paràmetres 0.01 i 0.01 com a anteriors. En el nostre cas, utilitzarem la distribució uniforme entre 0 i 10 que és la més senzilla i habitual.

## 5.2. Resultats del model espacial conjunt

Per elaborar el model conjunt dels tres càncers (còlon, estómac i esòfag) distingint per sexes, hem utilitzat la proposta de Gómez-Rubio et al. 2019 així com el codi que faciliten aquests autors en el seu article (vegeu Annex A2 per al cas conjunt de les dones).

Les figures 9 i 12 mostren la mitjana posterior de l'efecte espacial total, és a dir, la suma d'efectes compartits més els específics, per al cas de les dones i dels homes respectivament. La mitjana de la posteriori de l'efecte espacial compartit de la malaltia es mostren en les figures 11 i 14 , i les taules 4 i 5 indiquen els resums de la distribució a posteriori dels pesos de l'efecte espacial compartit per a cada malaltia, sempre distingint per sexes.

Comencem pel cas de les dones. Observem un patró molt similar entre el càncer d'estómac i d'esòfag tant a la Figura 9 com a la 10, amb una zona nord amb molta incidència i una zona sud amb poca. També s'observen valors més extrems en el cas del càncer d'estómac, però amb una diferència molt menys important que amb l'anàlisi conjunta.

Pel que fa als pesos, en el component espacial compartit, els càncers d'estómac i d'esòfag semblen tenir un pes molt similar i clarament superior al de còlon, confirmant que el càncer d'estómac i d'esòfag tenen una major dependència de l'efecte espacial compartit, és a dir, com hem vist, que el patró espacial és molt similar al patró compartit. En el cas del càncer de còlon, el pes petit indica poca dependència de l'efecte espacial compartit.

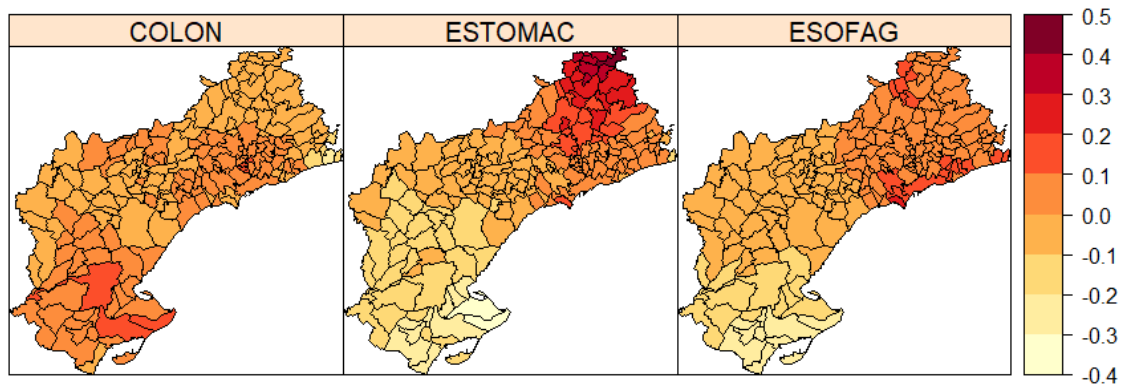


Figura 9. Posteriori de l'efecte espacial + efecte espacial específic de les dones.  
Font: elaboració pròpia

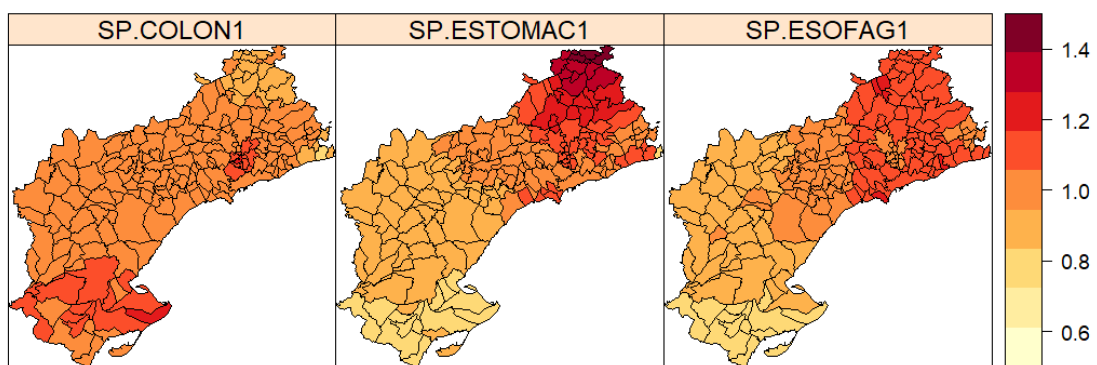


Figura 10. Posteriori de l'efecte espacial de les dones. Font: elaboració pròpia

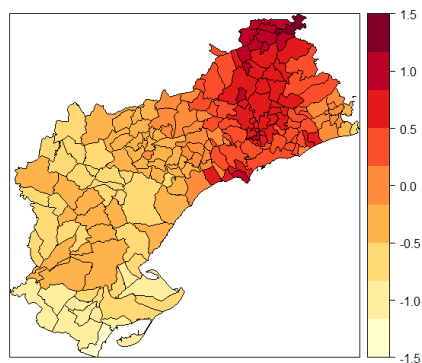


Figura 11. Efecte espacial compartit de les dones. Font: elaboració pròpia

Taula 4. Resum de les betes del model per a les dones. Font: elaboració pròpia

Beta de	Mínim	Primer quantil	Mediana	Tercer quantil	Màxim
Còlon	0,002	0,023	0,064	0,279	2,268
Estómac	0,009	0,055	0,247	0,757	8,547
Esòfag	0,015	0,113	0,289	0,868	11,778

El cas dels homes és bastant similar al de les dones. En la Figura 12 continuem veient el patró compartit entre el càncer d'estómac i d'esòfag, però, en aquest cas el patró de l'esòfag no és tan marcat, té uns valors menys extrems. En el cas del càncer d'estómac el comportament és pràcticament idèntic. Pel cas del càncer de còlon, el patró és molt similar al cas de les dones, però els valors són molt més extrems, sobretot en la part de Reus i Tarragona i, cap al sud, a Roquetes i Tortosa.

Pel que fa als pesos, tots tres tenen un pes molt similar, fet que es diferencia de les dones. Observem en la taula 4 que els pesos són pràcticament idèntics pels tres càncers. Per tant, els tres tipus de càncer tenen la mateixa dependència del component espacial compartit. Però cal comentar que aquests pesos són inferiors als d'estómac i d'esòfag del cas de les dones, és a dir, són molt similars però alhora també són menys dependents de l'efecte compartit.

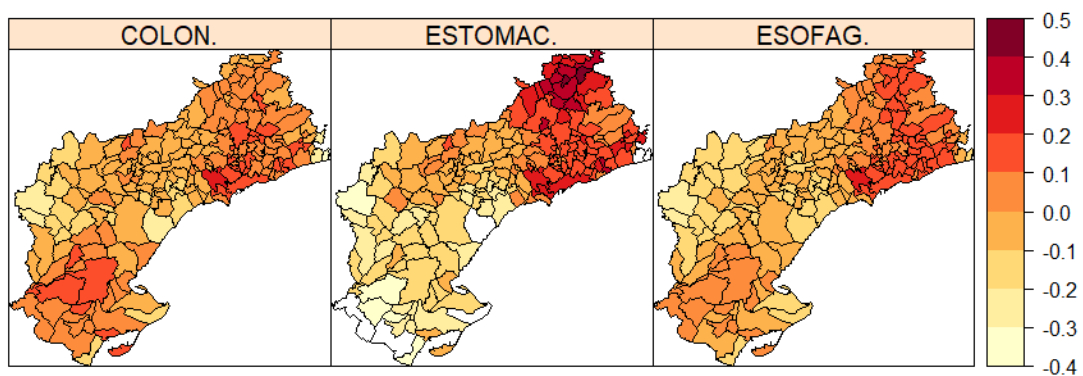


Figura 12. Posteriori de l'efecte espacial + efecte espacial específic dels homes.  
Font: elaboració pròpia

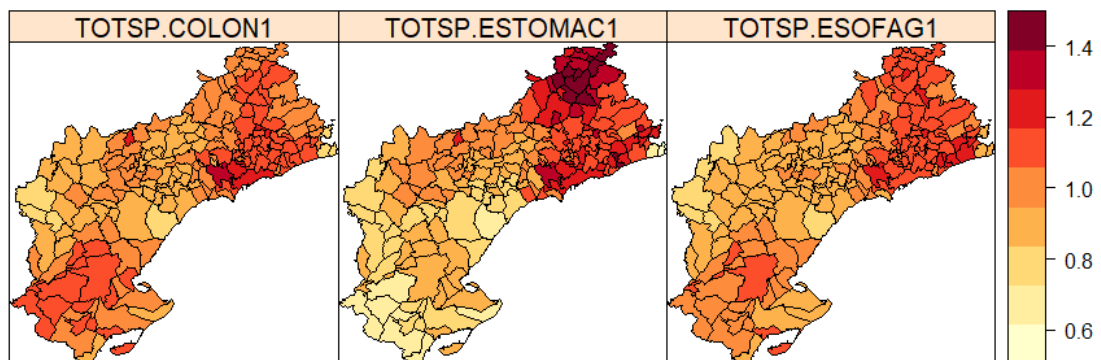


Figura 13. Posteriori de l'efecte espacial dels homes. Font: elaboració pròpia



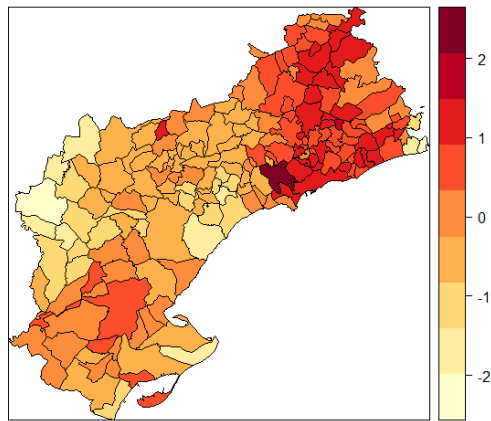


Figura 14. Efecte espacial compartit dels homes. Font: elaboració pròpia

Taula 5. Resum de les betes del model per als homes. Font: elaboració pròpia

Beta de	Mínim	Primer quantil	Mediana	Tercer quantil	Màxim
Còlon	0,028	0,057	0,162	0,522	3,311
Estómac	0,0224	0,067	0,173	0,467	2,692
Esòfag	0,021	0,051	0,168	0,455	2,568

## 6. Conclusions

Pel que fa al model individual, hem observat:

- El càncer de còlon té una clara major incidència en les zones més poblades, confirmant que la vida sedentària i d'hàbits urbans té una influència negativa. La distribució de les dones és més homogènia en el territori al voltant del valor 1, en els homes hi ha més variabilitat i en general la taxa d'incidència és una mica més baixa.
- En el càncer d'estómac la distribució de les taxes d'incidència és força similar entre homes i dones. La zona nord té una taxa d'incidència molt alta, sobretot en comparació amb la zona sud.
- El càncer d'esòfag té una taxa d'incidència més homogènia que els altres casos. La dels homes és més alta que la de les dones. En el cas de les dones la zona nord presenta una taxa una mica més alta que la resta.
- Si comparem els tres càncers entre ells no es veu un patró clar. Es podria considerar més similitud entre estómac i esòfag per al cas de les dones sempre tenint en compte la incidència més baixa de l'esòfag.
- En general, el comportament d'esòfag i estómac és d'alta incidència a la zona nord i baixa al sud, amb les ciutats grans com Reus i Tarragona amb altes taxes d'incidència.

Pel que fa al model conjunt, hem observat:

- El patró comú tant pels homes com per les dones marca clarament la zona nord amb major risc d'incidència. En el cas dels homes s'accentuen més les zones poblades i urbanes.
- Ara sí que observem major similitud, en el cas dels homes, entre la distribució dels càncers d'estómac i d'esòfag. Hi trobem més risc en la zona nord que en la sud, sobretot en el cas de l'estómac. També s'observa més clarament un risc elevat en les zones poblades pel que fa al càncer de còlon, en especial en el cas dels homes, això ens permet veure millor el seu comportament.
- Pel que fa als pesos, ens han confirmat la dependència compartida de l'estómac i l'esòfag cap al patró comú, ja que els dos tenien valors quasi idèntics. Els pesos dels homes són molt similars en les tres malalties, la qual cosa ens indica que les tres tenen la mateixa dependència cap al component comú. Cal dir que, els tres pesos en el cas dels homes són força inferiors als pesos de l'estómac i esòfag de les dones: homes al voltant de 0,16 i dones 0,25 per estómac i esòfag. Per tant, podem concloure que hem obtingut una forta relació entre estómac i esòfag de les dones, seguit de les tres malalties del cas dels homes, tal com s'ha apreciat en els gràfics.

Els dos models espacials utilitzats, individual i conjunt, ens han permès identificar patrons de comportament de les taxes d'incidència dels càncers de còlon, estómac i esòfag per homes i dones dels municipis de la província de Tarragona. Aquest tipus de resultats haurien de facilitar la presa de decisions tant en la prevenció com en el tractament d'aquestes malalties.

## 7. Bibliografia

- Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1), 1-20.
- Botella-Rocamora, P., Martínez-Beneito, M. A., & Banerjee, S. (2015). A unifying modeling framework for highly multivariate disease mapping. *Statistics in Medicine*, 34(9), 1548-1559.
- Downing, A., Forman, D., Gilthorpe, M. S., Edwards, K. L., & Manda, S. O. (2008). Joint disease mapping using six cancers in the Yorkshire region of England. *International Journal of Health Geographics*, 7(1), 1-14.
- Glaceran, J., Llop, D., Ameijide, A., Carulla, M., Miracle, C. (2019). *El mapa del càncer a Tarragona, 2000-2014*. Registre de Càncer de Tarragona, Servei d'Epidemiologia i Prevenció del Càncer, Hospital Universitari Sant Joan de Reus.
- Gómez-Rubio, V., Palmí-Perales, F., López-Abente, G., Ramis-Prieto, R., & Fernández-Navarro, P. (2019). Bayesian joint spatio-temporal analysis of multiple diseases. *SORT-Statistics and Operations Research Transactions*, 51-74.
- Knorr-Held, L., & Best, N. G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1), 73-85.
- Navarro Esteban, P. (2012). *Análisis espacial de riesgos de mortalidad por cáncer de próstata con INLA*. Trabajo Final de Máster, Universidad de Zaragoza.
- Otiende, V. A., Achia, T. N., & Mwambi, H. G. (2020). Bayesian hierarchical modeling of joint spatiotemporal risk patterns for Human Immunodeficiency Virus (HIV) and Tuberculosis (TB) in Kenya. *PloS one*, 15(7), e0234456.
- Província de Tarragona. (2022, 8 de març). Viquipèdia, l'Enciclopèdia Lliure. Data de consulta: 8/3/2022.  
[//ca.wikipedia.org/w/index.php?title=Prov%C3%ADncia\\_de\\_Tarragona&oldid=29619456](https://ca.wikipedia.org/w/index.php?title=Prov%C3%ADncia_de_Tarragona&oldid=29619456)
- Richardson, S., Abellan, J. J., & Best, N. (2006). Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK). *Statistical methods in medical research*, 15(4), 385-407.

## Annex. Codis d'R-INLA

### A1. Codi del model BYM per al cas individual còlon i dones

```
`` `{r,results='hide', warning=FALSE}
library(INLA)
library(spdep)
library(spacetime)
library(dplyr)
library(maptools)
library(readr)
library(readxl)
library(spdep)
library(sp)
library(RColorBrewer)
#load("")
...

`` `{r,warning=FALSE}
d<- read_excel("C:/Users/FelixBoschBosch/Desktop/tfg/ddones.xlsx")
#agafem el mapa
carto <- readShapePoly("C:/Users/FelixBoschBosch/Desktop/tfg/Dades TFG/Shape/tarragona.shp")
x.nb <- poly2nb(carto)
O <- d$Observed
E <- d$Expected
...

MAPES E,O i RIE
COMENTAR XAVI
`` `{r}
Ocol <- O[1:183]
Ecol <- E[1:183]
...

`` `{r}
#ptar1RIE <- SpatialPolygonsDataFrame(carto,
# data.frame(RIE = RIE),
# match.ID = FALSE )
#spplot(ptar1RIE,
# at = seq(-0.2, 2, by = 0.2), col.regions = gray(10:1/11))
...

Model INLA
`` `{r,warning=FALSE}
region <- 1:length(x.nb) #numero de municipi
dd <- data.frame(region=region, region.struct=region, O=Ocol, E=Ecol)
#creem base de dades per evitar utilitzar 'region' dues vegades
...

`` `{r}
#apliquem la funció dels veins
AdySP <- poly2nb(carto)
# creem la matriu d'incidencia
W.sp <- as(nb2mat(AdySP, style = "B"), "Matrix")
```

```

...
``{r}
#creem formula
formula <- Ocol ~ f(region.struct,model="besag",graph=W.sp,
param=c(1,0.001))+f(region,model="iid", param=c(1,0.01))
#apliquem INLA
p1<-inla(formula, family="poisson", data=dd, E=Ecol,
control.compute=list(dic=T, cpo=TRUE),
control.predictor=list(compute=TRUE,cdf=c(log(1))),
control.inla=list(strategy="laplace",int.strategy="grid"))
...
``{r}
#preperem les variables pel gràfic, la posterior mean
ptar1 <- SpatialPolygonsDataFrame(carto,
  data.frame(_SHARED = p1$summary.fitted.values[, "mean"]),
  match.ID = FALSE )
ptar1$SP.COLON <- p1$summary.fitted.values[, "mean"]
...
``{r}
#Fem el gràfic
my.palette1 <- brewer.pal(n = 9, name = "YlOrRd")
spplot(ptar1, c("SP.COLON"),
  at = seq(0.74, 1.23, by = 0.49/9), col.regions = my.palette1, main="Còlon Dones")
...

```

## A2. Codi del model BYM per al cas conjunt dones

```
``{r,results='hide'}
library(INLA)
library(spdep)
library(spacetime)
library(dplyr)
library(maptools)
library(readr)
library(readxl)
library(spdep)
library(RColorBrewer)
``
``{r}
d<- read_excel("C:/Users/FelixBoschBosch/Desktop/tfg/ddones.xlsx")

map <- readShapePoly("C:/Users/FelixBoschBosch/Desktop/tfg/Dades TFG/Shape/tarragona.shp")

n.st <- nrow(map)

d$sp.idx <- d$CPROV

d$disease <- rep(1:3, each = n.st)
d$intercept <- as.factor(d$disease)

d$s.dummy <- NA

d$s.1 <- NA
d$s.1 [d$disease == 1] <- d$sp.idx [d$disease == 1]
d$s.2 <- NA
d$s.2 [d$disease == 2] <- d$sp.idx [d$disease == 2]
d$s.3 <- NA
d$s.3 [d$disease == 3] <- d$sp.idx [d$disease == 3]

AdySP <- poly2nb(map)
W.sp <- as(nb2mat(AdySP, style = "B"), "Matrix")
prior.beta.s = list(prior = "normal", param = c(0, 1 / 5.9),
                    fixed = FALSE, initial = 0.01)

file.saved <- "st_model_UNIFORM.RData"
prior.prec = list(prior =
                  "expression:
logdens = -log_precision / 2;
return(logdens)",
                  initial = 0)
d$sp.idx1 <- d$s.1
d$sp.idx2 <- d$s.2
d$sp.idx3 <- d$s.3
``
``{r}
inla.scale <- FALSE
formula = Observed ~ -1 +
intercept +
```

```

f(sp.idx1, model = "besag", scale.model = inla.scale,
  graph = W.sp, hyper = list(prec = prior.prec)) +
f(sp.idx2, model = "besag", scale.model = inla.scale,
  graph = W.sp, hyper = list(prec = prior.prec)) +
f(sp.idx3, model = "besag", scale.model = inla.scale,
  graph = W.sp, hyper = list(prec = prior.prec)) +
f(s.dummy, model = "besag", scale.model = inla.scale, graph = W.sp,
  hyper = list(prec = prior.prec))+
f(s.1, copy = "s.dummy", range = c(0, Inf),
  hyper = list(beta = prior.beta.s)) +
f(s.2, copy = "s.dummy", range = c(0, Inf),
  hyper = list(beta = prior.beta.s)) +
f(s.3, copy = "s.dummy", range = c(0, Inf),
  hyper = list(beta = prior.beta.s))
...
````{r, results='hide'}```
r = inla(formula, data = d, family = "poisson", control.predictor=list(compute=TRUE),
  E = Expected, verbose = TRUE,
  control.compute = list(config = TRUE),
  control.inla(strategy = "laplace", npoints = 21))

r <- inla.rerun(r)
...
````{r}```
summary(r)
...
````{r}```
tar <- SpatialPolygonsDataFrame(map,
  data.frame(_SHARED = r$summary.random$s.dummy[, "mean"]),
  match.ID = FALSE )
...
````{r}```
tar$COLON <- r$summary.random$sp.idx1[, "mean"]
tar$ESTOMAC <- r$summary.random$sp.idx2[, "mean"]
tar$ESOFAG <- r$summary.random$sp.idx3[, "mean"]

my.palette1 <- brewer.pal(n = 9, name = "YlOrRd")
...
````{r}```
spplot(tar, c("COLON", "ESTOMAC", "ESOFAG"),
  at = seq(-0.41, 0.49, by = 0.1), col.regions = my.palette1, main="Posteriori de l'efecte espacial")

spplot(tar, c("SHARED"),
  at = seq(-2.57, 2.64, by = 5.21/9), col.regions = my.palette1, main='Efecte espacial compartit')
...
````{r}```
tar$COLON. <- r$summary.random$sp.idx1[, "mean"] +
  r$summary.random$s.1[, "mean"]
tar$ESTOMAC. <- r$summary.random$sp.idx2[, "mean"] +
  r$summary.random$s.2[, "mean"]
tar$ESOFAG. <- r$summary.random$sp.idx3[, "mean"] +
  r$summary.random$s.3[, "mean"]

```

```

...
``{r}
splot(tar, c("COLON.", "ESTOMAC.", "ESOFAG."),
      at = seq(-0.4, 0.5, by = 0.1), col.regions = my.palette1, main = "Posteriori de l'efecte espacial +
l'efecte espacial específic de cada malaltia")
...
``{r}
samples <- inla.posterior.sample(1000, r)
betas <- sapply(1:1000, function(x) (samples[[x]]$hyperpar)[5:7])
betas <- as.data.frame(t(betas))
plot(betas)
cor(betas)
...
``{r}
summary(betas)

xx <- (r$joint.hyper)
probs <- xx[, 8]
probs <- exp(probs - max(probs))
probs <- probs / sum(probs)
summary(xx)
...

```