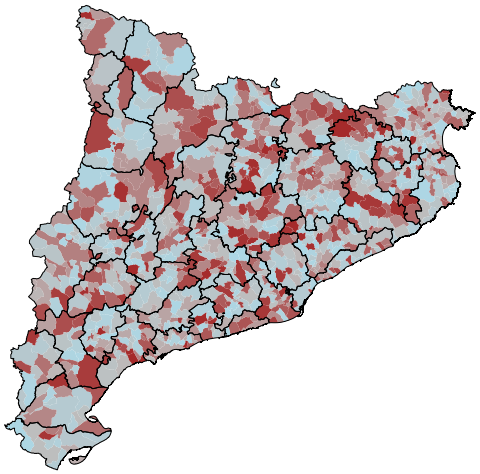
DETECTION OF GEOGRAPHICAL CLUSTERS OF PREMATURITY IN CATALONIA

Joe Brew

Erasmus Mundus, FetalMed PhD Proposal



Project summary

Though clinical treatments for infants born prematurely and/or at low weight have improved, there is an urgent need for research into these adverse events' causes. Of particular need is a framework for identifying and understanding prematurity and low weight birth's spatial components, so as to direct preventive health interventions. Using spatial statistics and novel machine learning methods, this project aims to develop a framework through which Catalonia can identify clusters indicative of adverse prenatal health exposures, and then expand that framework to other European regions.

• joebrew@gmail.com •

General information

Project:
Detection of geographical clusters of prematurity in Catalonia

Investigators:
Joe Brew, MPH, MA (applicant)
Montse Palacio, MD (requested supervisor)

Research questions:
What factors affect the likelihood of premature birth in Catalonia? After adjustment for relevant sociodemographic confounders, can geographic information system (GIS) tools and spatial statistics be used to predict cases of prematurity? To what extent do environmental exposures explain spatially differential risk? How can municipal and regional public health authorities intervene to prevent adverse birth outcomes in high-risk areas? Can models and methods devised for prediction in Catalonia be applied in other European regions?

Importance / Rationale:
In the first moments of life, preterm and low birth weight (LBW) infants are at a greater risk of neonatal mortality [1, 2] and morbidity [3, 4]. When these infants survive, they face a higher likelihood of chronic conditions as adults [5, 6].

There is a growing body of research into the negative *consequences* of prematurity, and a greater push for newborn-specific interventions to reduce neonatal mortality. [7, 8] However, there is a lack of research into its *causes*, particularly in regards to prevention, [9, 10]. Social and demographic risk factors have been established [11, 12], but many of them are not "intervenible" [13, 14, 15]. Recent technological advances have enabled better clinical detection methods [16, 17], but detection does not, in itself, *prevent*.

Environmental exposures, particularly the inhalation of fine particulate matter and residing in areas with air pollution , have been identified as an important intervenible risk factor for prematurity and LBW both in Spain [18] and internationally [19, 20, 21]. However, most environmental studies are carried out at an aggregated level, and do not account for variation in level of exposure within cities or regions. These studies, though useful to national-level authorities, are of little value to local and regional-level public health practitioners. An understanding of environmental exposures at the local level is essential to primary prevention, and the first step in identifying these exposures is the identification of spatial clusters.

As with clinical screening methods, tools for epidemiological investigation and geospatial analysis have improved rapidly in recent years. This project combines needed research into an area of relative uncertainty (detection of intervenible public health threats to prenatal health and differential risk down to the neighborhood level) with modern tools (GIS) and methods (spatial scans, clustering detection, etc.), so as to better understand a problem of significant public health importance.

Study objectives

The objectives of this project are two-fold:

1. Identify both spatial and non-spatial risk factors for LBW and premature birth in the region of Catalonia using novel machine learning approaches and GIS tools, with a focus on prenatal environmental exposures at the neighborhood level.

2. Given the results of the first objective, test the model's predictive ability on non-Catalan data, and design an intervention framework for the prevention of LBW and prematurity for European cities and regions at large.

Study Design

Design: Prospective, observational cohort study

Research population: Expectant mothers in their first or second trimester residing in Catalonia, Spain during the study recruitment time period.

Time period: Study participants will be recruited from October, 2015 through March, 2016.

Methodology

Data acquisition and organization: Following approval from the appropriate ethical committees, institutional review boards, and governmental agencies, pregnant women in their first or second trimesters of pregnancy will be recruited to participate in the study. "Participation" constitutes (a) a simple 5-10 minute survey outlining basic socioeconomic, demographic and medical traits/history as well as residential, family and labor market information and (b) agreement to allow access to their birth outcome data (gestional age, weight at birth, survival and defects status).

Exclusion criteria: Given the simplicity of the survey and lack of any risk for participants, there are no explicit exclusion criteria in the data collection phase. In the analysis phase, subjects may be excluded given certain pre-existing health conditions or non-representative demographic traits.

Safety Considerations

There are no reasonable foreseeable safety considerations associated with this research.

Data Management and Statistical Analysis

Data management: Private health data will be stored on a password-protected and fully encrypted hard-drive, and will accessed locally only by the principal investigators. Following initial collection and cleaning, data will be "anonymized" by using unique key-pairs (ID numbers linked to mother's name), stored in a separate linkage document to which only the principal investigators have access.

Statistical analysis: The analysis phase of the project has three phases:

1. Feature generation, aggregation and public data joins
2. Model construction
3. Model testing

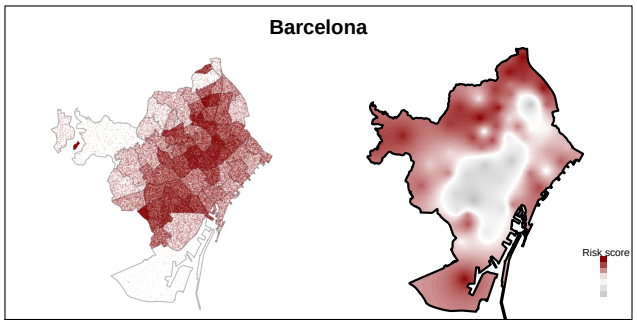
In phase 1, the dataset constructed through the surveying and birth outcomes collection will be joined to Spanish census data as well as relevant commercial and infrastructure data (proximity to major roads, industrial sites, etc.). Additionally, study participants will be stratified into risk "groups" based on established non-spatial risk factors for LBW and premature birth (age, nutrition factors, family-social environment, socioeconomic status, immigration status, etc.).

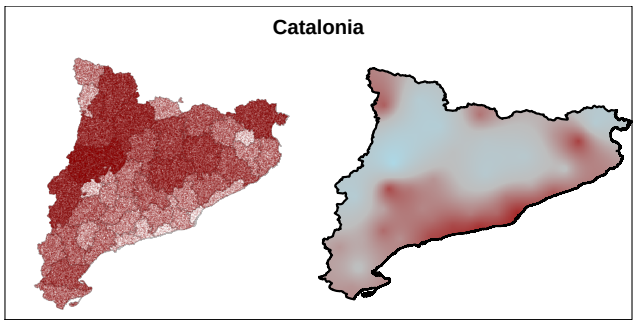
In phases 2 and 3, analysis will take a standard machine learning approach to variable/model selection and validation. Statistical models will be constructed with the aim of predicting the likelihood of prematurity given a combination of relevant spatial and non-spatial features. 6 different model types (logistic regression, generalized additive model, krieging, Kulldorff's spatial scan, random forest, and support vector regression) will be "trained" on a randomly selected subset (constituting two thirds of the total number of study participants), and then "tested" on the remaining third's outcomes. Model strengths, weaknesses and extrapolatability will be explored through standardized tests.

Data deliverable: The incidence of prematurity will be presented both by space as well as mother's characteristics. This deliverable will guide further research as well as inform Catalan public health authorities (who can identify neighborhoods of greatest risk, and therefore greatest need) as to who is most at risk and where.

In order to parse out potential environmental effects, the final model will be used to predict the risk of prematurity using a homogenous population. The maps on the following page provide examples of the outcome. By standardizing non-spatial features, risk by units as small as urban neighborhood (top left) as well as a spatial risk "surface" for entire cities (top right) can be constructed. Likewise, at the regional level, individual comarcas can compare their county-wide risk score (bottom left), whereas regional authorities can identify spatial-temporal patterns in prematurity (bottom right).[1] The population-standardized risk "surface" will be particularly useful in the detection of potential environmental exposures affecting fetal and maternal health.

[1] All maps constructed using entirely simulated data.





Research Deliverables: Dissemination of Results and Publication Policy

This study has both scientific (objective 1) and policy (objective 2) components. For the former, results will be published in academic journals, with credit attributed to the relevant universities, governmental agencies and Erasmus Mundus organization. For the latter, the main objective is the development and dissemination of a predictive framework applicable to (and useful for) other European regions. This "framework" consists of both a "toolkit" (open-source statistical code) as well as an interactive, multi-lingual training component (web-based) to help guide public health practitioners in the identification of LBW and prematurity clusters in their own cities and regions.

Expected Outcomes of the Study

Given the established importance of environmental factors to maternal and infant health outcomes, it is reasonable to foresee the identification of "clusters" which cannot be fully explained by their non-spatial components. The positive identification of both kinds of clusters (those which can be fully explained by the make-up of the residents, and those which appear to have a spatial / environmental exposure component) will be of use to public health practice and research.

Duration of the Project

In general terms, the first year of the project will consist of survey and study design refinement, the second year will consist of data cleaning, feature generation and analysis, and the third year will be devoted to publication, dissemination and the construction of the European regions' "toolkit."

Problems Anticipated

None.

Project Management

This project will be carried out in accordance with the regulations and guidelines of the Erasmus Mundus and FetalMed PhD programs, under the direction of Montse Palacio (the requested project director), with the principal research being carried out by Joe Brew (the PhD applicant).

Ethics

Mr. Brew will seek approval from the relevant committees, conforming both to the guidelines of law as well as best practice.

Informed consent

Informed consent will be provided in the form of a letter as well as the attending physician's explanation. The letter will be available in Catalan, Spanish, French, English, Portuguese, Arabic and Turkish. For potential participants who do not read, or who do not understand the aforementioned languages, a medical interpretor will be sought both for the purposes of the visit, as well as the translation of the letter of informed consent. Participants will be asked to sign to signify their agreement with the study terms. A participant may remove themselves from the study upon simple request at any point in the data collection or analysis phases. There are no incentives or de-incentives for study participation.

Budget

The budget for this project will be minimal, and incidental costs (printing, computing, website-hosting, etc.) will be incurred by Mr. Brew.

Other support for the Project

Mr. Brew seeks a Category A Erasmus Mundus grant for the purposes of this project.

Collaboration with other scientists or research institutions

Particularly in the accomplishment of objective 2 (the construction of a predictive framework for LBW and premature birth for other European regions), collaboration will be a necessary component of this project. Mr. Brew will rely on his large professional and academic network (he has previously studied at the Anadalusian School of Public Health in Spain, The Copenhagen School of Global Health in Denmark, and the School of Higher Studies in Public Health in France) for (a) advising and assistance, (b) identification of needs and principal stakeholders for the predictive framework and (c) "toolkit" testing and feedback. Mr. Brew also seeks to draw on Dr. Palacio's clinical expertise and network so as to fully utilize practitioners' local knowledge and experience in both model construction and validation.

Credentials of the investigators

Mr. Brew is an experienced data scientist and public health practitioner. He currently works as the surveillance epidemiologist for the Florida Department of Health in Alachua County. His work for FDOH has focused on low birth weight, infant mortality, and geography. He has previously worked as a "Data Science for Social Good" fellow with the Chicago Department of Public Health, where he developed a predictive model to identify lead poisoning cases among infants and children.

Dr. Palacio, Maternal-Fetal consultant and Chair of the Obstetric and High Risk Obstetrics wards at the Hospital Universitari, is an expert in the field of fetal medicine. Her work in the 'Unitat de Prematuritat', and her combination of subject matter and clinical experience, as well as her familiarity with Barcelona's clinical and data systems, are essential to this project's success.

Format

This protocol follows the format recommended by the World Health Organization.

Contact

Joe Brew, MPH, MA
530 NW 2nd Street
Gainesville, FL 32601
USA
352.318.4553
joebrew@gmail.com

References

[1] Li-Yi Tsai, Yi-Ling Chen, Kuo-Inn Tsou, and Shu-Chi Mu. The impact of small for gestational age on neonatal outcome among very low birth weight infants. *Pediatrics & Neonatology*, oct 2014. doi: 10.1016/j.pedneo.2014.07.007. URL http://dx.doi.org/10.1016/j.pedneo.2014.07.007.

[2] Anna W. Anderson, P. Brian Smith, Kristin M. Corey, Kevin D. Hill, Kanecia O. Zimmerman, Reese H. Clark, and Christoph P. Hornik. Clinical outcomes in very low birth weight infants with major congenital heart defects. *Early Human Development*, 90(12):791–795, dec 2014. doi: 10.1016/j.earlhumdev.2014.09.006. URL http://dx.doi.org/10.1016/j.earlhumdev.2014.09.006.

[3] Edwina H. Yeung, Candace Robledo, Nansi Boghossian, Cuilin Zhang, and Pauline Mendola. Developmental origins of cardiovascular disease. *Curr Epidemiol Rep*, 1(1):9–16, jan 2014. doi: 10.1007/s40471-014-0006-4. URL http://dx.doi.org/10.1007/s40471-014-0006-4.

[4] Myrte Merkestein, Felino R. Cagampang, and Dyan Sellayah. Fetal programming of adipose tissue function: an evolutionary perspective. *Mammalian Genome*, 25(9-10):413–423, jun 2014. doi: 10.1007/s00335-014-9528-9. URL http://dx.doi.org/10.1007/s00335-014-9528-9.

[5] Silvia Visentin, Francesca Grumolato, Giovanni Battista Nardelli, Barbara Di Camillo, Enrico Grisan, and Erich Cosmi. Early origins of adult disease: Low birth weight and vascular remodeling. *Atherosclerosis*, 237(2):391–399, dec 2014. doi: 10.1016/j.atherosclerosis.2014.09.027. URL http://dx.doi.org/10.1016/j.atherosclerosis.2014.09.027.

[6] Robert H. Lane. Fetal programming, epigenetics, and adult onset disease. *Clinics in Perinatology*, 41(4):815–831, dec 2014. doi: 10.1016/j.clp.2014.08.006. URL http://dx.doi.org/10.1016/j.clp.2014.08.006.

[7] Eric C. Eichenwald and Ann R. Stark. Management and outcomes of very low birth weight. *New England Journal of Medicine*, 358(16):1700–1711, apr 2008. doi: 10.1056/nejmra0707601. URL http://dx.doi.org/10.1056/NEJMra0707601.

[8] Tessa Wardlaw, Danzhen You, Lucia Hug, Agbessi Amouzou, and Holly Newby. UNICEF report: enormous progress in child survival but greater focus on newborns urgently needed. *Reproductive Health*, 11(1):82, 2014. doi: 10.1186/1742-4755-11-82. URL http://dx.doi.org/10.1186/1742-4755-11-82.

[9] Sorina Grisaru-Granovsky, Brian Reichman, Liat Lerner-Geva, Valentina Boyko, Cathy Hammerman, Arnon Samueloff, and Michael S. Schimmel. Population-based trends in mortality and neonatal morbidities among singleton, very preterm, very low birth weight infants over 16 years. *Early Human Development*, 90(12):821–827, dec 2014. doi: 10.1016/j.earlhumdev.2014.08.009. URL http://dx.doi.org/10.1016/j.earlhumdev.2014.08.009.

[10] C. E. Rubens, Y. Sadovsky, L. Muglia, M. G. Gravett, E. Lackritz, and C. Gravett. Prevention of preterm birth: Harnessing science to address the global epidemic. *Science Translational Medicine*, 6(262):262sr5–262sr5, nov 2014. doi: 10.1126/scitranslmed.3009871. URL http://dx.doi.org/10.1126/scitranslmed.3009871.

[11] Natalie Slopen, Eric B. Loucks, Allison A. Appleton, Ichiro Kawachi, Laura D. Kubzansky, Amy L. Non, Stephen Buka, and Stephen E. Gilman. Early origins of inflammation: An examination of prenatal and childhood social adversity in a prospective cohort study. *Psychoneuroendocrinology*, 51:403–413, jan 2015. doi: 10.1016/j.psyneuen.2014.10.016. URL http://dx.doi.org/10.1016/j.psyneuen.2014.10.016.

[12] Faith H Yego, Catherine D'Este, Julie Byles, Paul Nyongesa, and Jennifer Stewart Williams. A case-control study of risk factors for fetal and early neonatal deaths in a tertiary hospital in kenya. *BMC Pregnancy and Childbirth*, 14(1), nov 2014. doi: 10.1186/s12884-014-0389-8. URL http://dx.doi.org/10.1186/s12884-014-0389-8.

[13] Ashley H. Schempf, Amy M. Branum, Susan L. Lukacs, and Kenneth C. Schoendorf. Maternal age and parity-associated risks of preterm birth: differences by race/ethnicity. *Paediatric and Perinatal Epidemiology*, 21(1):34–43, jan 2007. doi: 10.1111/j.1365-3016.2007.00785.x. URL http://dx.doi.org/10.1111/j.1365-3016.2007.00785.x.

[14] Prakesh S. Shah. Parity and low birth weight and preterm birth: a systematic review and meta-analyses. *Acta Obstetricia et Gynecologica Scandinavica*, 89(7):862–875, jul 2010. doi: 10.3109/00016349.2010.486827. URL http://dx.doi.org/10.3109/00016349.2010.486827.

[15] Naoko Kozuki, Anne CC Lee, Mariangela F Silveira, Ayesha Sania, Joshua P Vogel, Linda Adair, Fernando Barros, Laura E Caulfield, Parul Christian, Wafaie Fawzi, Jean Humphrey, Lieven Huybregts, Aroonsri Mongkolchati, Robert Ntozini, David Osrin, Dominique Roberfroid, James Tielsch, Anjana Vaidya, Robert E Black, and Joanne Katz. The associations of parity and maternal age with small-for-gestational-age, preterm, and neonatal and infant mortality: a meta-analysis. *BMC Public Health*, 13(Suppl 3):S2, 2013. doi: 10.1186/1471-2458-13-s3-s2. URL http://dx.doi.org/10.1186/1471-2458-13-S3-S2.

[16] Alan Roberto Hatanaka, Rosiane Mattar, Tatiana Emy Nishimoto Kawanami, Marcelo Santucci França, Liliam Cristine Rolo, Roseli Mieko Yamamoto Nomura, Edward Araujo Júnior, Luciano Marcondes Machado Nardozza, and Antonio Fernandes Moron. Amniotic fluid 'sludge' is an independent risk factor for preterm delivery. *J Matern Fetal Neonatal Med*, pages 1–20, nov 2014. doi: 10.3109/14767058.2014.989202. URL http://dx.doi.org/10.3109/14767058.2014.989202.

[17] Ronald J. Wapner, Joshua E. Babiarz, Brynn Levy, Melissa Stosic, Bernhard Zimmermann, Styrmir Sigurjonsson, Nicholas Wayham, Allison Ryan, Milena Banjevic, Phil Lacroute, Jing Hu, Megan P. Hall, Zachary Demko, Asim Siddiqui, Matthew Rabinowitz, Susan J. Gross, Matthew Hill, and Peter Benn. Expanding the scope of non-invasive prenatal testing: Detection of fetal microdeletion syndromes. *American Journal of Obstetrics and Gynecology*, dec 2014. doi: 10.1016/j.ajog.2014.11.041. URL http://dx.doi.org/10.1016/j.ajog.2014.11.041.

[18] Exposure to traffic pollution before birth could impair lung function. *Nursing Standard*, 29(14):16–16, dec 2014. doi: 10.7748/ns.29.14.16.s19. URL http://dx.doi.org/10.7748/ns.29.14.16.s19.

[19] Takashi Yorifuji, Hiroo Naruse, Saori Kashima, Takeshi Murakoshi, and Hiroyuki Doi. Residential proximity to major roads and obstetrical complications. *Science of The Total Environment*, 508:188–192, mar 2015. doi: 10.1016/j.scitotenv.2014.11.077. URL http://dx.doi.org/10.1016/j.scitotenv.2014.11.077.

[20] E. Morales, R. Garcia-Esteban, O. Asensio de la Cruz, M. Basterrechea, A. Lertxundi, M. D. Martinez Lopez de Dicastillo, C. Zabaleta, and J. Sunyer. Intrauterine and early postnatal exposure to outdoor air pollution and lung function at preschool age. *Thorax*, oct 2014. doi: 10.1136/thoraxjnl-2014-205413. URL http://dx.doi.org/10.1136/thoraxjnl-2014-205413.

[21] Juliana Oliveira de Melo, Sônia Fátima Soto, Isis Akemi Katayama, Camilla Ferreira Wenceslau, Amanda Gonçalves Pires, Mariana Matera Veras, Luzia N.S. Furukawa, Isac de Castro, Paulo Hilário Nascimento Saldiva, and Joel Claudio Heimann. Inhalation of fine particulate matter during pregnancy increased IL-4 cytokine levels in the fetal portion of the placenta. *Toxicology Letters*, dec 2014. doi: 10.1016/j.toxlet.2014.12.001. URL http://dx.doi.org/10.1016/j.toxlet.2014.12.001.