

Functional response regression with funBART: an analysis of patient-specific stillbirth risk:

Data Dictionary for reconstructing stillbirth dataset from National Center for Health Statistics publicly available files.

Introduction:

Our dataset includes data files publicly available from the National Center for Health Statistics. It contains Linked Birth/Infant Death Cohort Data files (2004 – 2006) and Fetal Death Data files (2004 – 2006). We limit our analysis to complete cases for pregnancies delivered from 34 to 42 weeks' gestation.

External Link:

Data files and the CDC data dictionary are publicly available at https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Period_Linked

New variables:

New variables were created that contained information for infant death, abortion, neonatal death, fetal death, and live birth. Details are in following section. A new variable was created to denote the year the data corresponds to.

Our Dataset:

Variable	Description
nd	<p>Binary indicator for neonatal death (within 28 days after delivery). This variable is not present in the original dataset; it was created as follows.</p> <pre>nd = 1 if{ death_age_days<=28 and livebirth==1 }</pre>
fd	<p>Binary indicator for fetal death (stillbirth; death in utero > 20 weeks gestation). This variable is not present in the original dataset; it was created as follows.</p> <pre>fd = 1 if{ fetal_death_year !=. and gestation_detail>0 }</pre> <p>We note that we use this variable as an indicator for stillbirth, defined as death in utero after 20 weeks' gestation, as we are limiting our dataset to 34–42 weeks' gestation. Were this not the case, we could create a separate stillbirth variable as follows:</p> <pre>sb =1 if{ fetal_death_year !=. and gestation_detail >= 20 }</pre>

	}
livebirth	Binary indicator for live birth. This variable is not present in the original dataset; it was created as follows. <pre>livebirth = 1 if{ fetal_death==0 and gestation_detail != . }</pre>
gest_age	gestational age of delivery, in weeks. CDC variable combgest.
induce	Binary indicator for inducement of labor. CDC variable uop_induc.
hirisk	Binary indicator for high risk pregnancy, based on presence of at least one of: anemia, cardiac, diabetes, hemoglob, htn, lung, renal, rhsens. Set to 1 if at least one of these variables is 1; else 0.
priorLive	Binary indicator for number of prior live births > 0. Uses CDC variable priorlive.
priorDead	Binary indicator for number of prior births now dead > 0. Uses CDC variable priordead.
priorTerms	Binary indicator for number of prior other terminations > 0. Uses CDC variable priorterm.
wtGain	Maternal weight gain during pregnancy. CDC variable wtgain.
momAge	Maternal age in years. CDC variable mager.
agegrp	Discretizes the momAge variable into categories: <20, 20-24, 25-29, 30-34, 35-39, 40-44, and 45+.
momEthnicity	Maternal ethnicity: White non-hispanic, Black non-hispanic, Hispanic, and Other. CDC variable mracehisp. If mracehisp in 1-5, "Hispanic". If mracehisp == 6, "White non-hispanic". If mracehisp == 7, "Black non-hispanic". Else "Other."
anemia	Indicator for presence of maternal anemia. CDC variable urf_anemia,
cardiac	Indicator for presence of maternal cardiac disease. CDC variable urf_card.
diabetes	Indicator for presence of maternal diabetes (excluding gestational). CDC variable urf_diab.
hemoglob	Indicator for presence of maternal hemoglobinopathy. CDC variable urf_hemo.
htn	Indicator for presence of maternal chronic hypertension. CDC variable urf_chyper.
lung	Indicator for presence of maternal lung disease. CDC variable urf_lung.
renal	Indicator for presence of maternal renal disease. CDC variable urf_renal.
rhsens	Indicator for presence of maternal Rh sensitization. CDC variable urf_rh.
multiparous	Indicator for multiparity. Based on CDC variable
male	Indicator for male infant sex. CDC variable sex.
birthwt	Infant birth weight, in grams. CDC variable dbwt.
year	Year of delivery. This variable was created to indicate which year of data each record is sourced from.
gest_age34	Gestational age, with 34 weeks normalized to 0.
otherRisk	Indicator for all conditions included in hirisk, excluding diabetes and htn.
diab_htn	Diabetes and hypertension maternal status (Neither, Diabetes, Hypertension, or Both). CDC variable names urf_chyper, urf_diab.

id	Unique identifier for each pregnancy, based on row numbers of the data.
wtgainQ	Maternal weight gain quantile, based on momAge: [0.00,0.10), [0.10,0.25), [0.25,0.75), [0.75,0.90), [0.90,1.00]
wtgainQ_vals	Corresponding weight gain ranges for each weight gain quantile range.
birthwtQ	Birth weight quantile, based on birthwt: [0.00,0.10), [0.10,0.25), [0.25,0.75), [0.75,0.90), [0.90,1.00]
birthwtQ_vals	Corresponding birth weight ranges for each birth weight quantile range.
testtrain	'test' or 'train', may be used for a stratified-sampling test/train split.
source	data and dictionary are available at https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm