

PIN STUDY – COHORTS 1 & 2
Study-wide analysis variables - specifications for construction

dataset stored in: \prom\data\sas\study\studyvar.x__
program to generate in: \prom\data\programs\study\studyvar.pgm

	type*	source	source var name	constructed var name	constructed variable values
study participant ID – required to merge with other datasets				PATID	
Demographics					
mom's race **NOTE: source files had inconsistent coding; this coding from the Main Q	H	main Q tracking file	mqq7; momrace	c_mrace	1=white 2=African American 3=American Indian 4=Asian or Pacific Islander <i>see attached list for other codes</i>
mom's ethnicity	H	main Q tracking file	mqq6; momrace (hispanic code)	c_methni	0=not hispanic 1=hispanic
mom's marital status **NOTE: source files had inconsistent coding; this coding from the Main Q	H	main Q tracking file	mqq4; marital	c_marita	1=single 2=married 3=separated 4=divorced 5=widowed
mom's age at 24 weeks' gestation	H	main Q tracking file	mqq1a, mqq1b, mqq1c;	c_mage24	
mom's highest level of educ.	H	main Q tracking file	mqq10; momeduc	c_momedu	grade level completed

* H=hierarchical, C=constructed
V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

	type*	source	source var name	constructed var name	constructed variable values
mom's household % of poverty for 1996 <u>source:</u> <i>U.S. Bureau of the Census, Current Population Reports, Series P60-198, Poverty in the United States: 1996, U.S. Government Printing Office, Washington, DC, 1997, p. 1.</i>	C	main Q	mqq14, mqq15, mqq16	c_povert	value of 100 is at the poverty level, based on income, # of adults, & # of children in h-hold
Mom's height, weight, and BMI					
mom's height	H	screeener	cs9a, cs9b	c_inches	
pre-pregnancy weight – all sources		screeener med abs		c_prewt	
mom's pre-pregnancy weight OLD - DO NOT USE	H	screeener	cs10, (impute from: csgestwt, csgestdt)	c_lbs	
mom's pre-pregnancy BMI (kg/m ²) – some values are imputed	C	screeener		c_bmi	
mom's pre-pregnancy BMI (categorized) Based on IOM guidelines	C	med abs screeener		c_bmiiom	1= BMI < 19.8 -- low 2= 19.8 <= BMI < =26.0 -- normal 3= 26.0 < BMI < =29.0 -- overweight 4= BMI >29.0 -- obese
Current pregnancy info					
OLD ALGORITHM indicates preterm case classification was based on delivery prior to 37 weeks' completed gestation, using an algorithm of reported last menstrual period (LMP) and/or first ultrasound performed prior to recruitment. Gestational age was assessed based on a reliable estimate of LMP, when available, or first ultrasound if LMP was unknown. When both were available and the two estimates were within 14 days of one another, the last menstrual period date was used. When the disagreement exceeded 14 days, the ultrasound estimate was used.					
NEW ALGORITHM indicates preterm case classification was based on delivery prior to 37 weeks' completed gestation, using an algorithm based on the first ultrasound performed prior to 22 weeks' gestation (up to 21 weeks, 6 days). If no ultrasound was performed or none was performed prior to the					

* H=hierarchical, C=constructed

V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

	type*	source	source var name	constructed var name	constructed variable values		
start of week 22, then LMP was used to date the pregnancy.							
BOTH ALGORITHMS – The resulting “start of pregnancy” (actually an LMP-equivalent) based on the LMP or U/S described above was used to anchor all gestational age calculations for project activities, including definition of the recruitment window, follow-up dates, and ultimately preterm-term outcome.							
		Bestedc - based on priority of ultrasound estimated edc over Imp date NEW ALGORITHM DATE VARIABLE – NOT RELEASED	C	tracking	ultraedc eddlmp	bestedc2 (new)	SAS date format
		Start of pregnancy date (based on bestedc2) NEW ALGORITHM DATE VARIABLE – NOT RELEASED	C	tracking	bestedc2	strtprg2 (new)	SAS date format
		gestational age at delivery, weeks OLD ALGORITHM	C	tracking	delivdat, startprg	c_delges (old)	completed weeks (based on old algorithm)
		gestational age at delivery, weeks NEW ALGORITHM	C	tracking	delivdat, strtprg2	c_delgs2 (new)	completed weeks (based on new algorithm)
		gestational age at delivery, days OLD ALGORITHM		tracking	delivdat, startprg	c_gsday	(based on old algorithm)
		gestational age at delivery, days NEW ALGORITHM		tracking	delivdat, strtprg2	c_gsday2	(based on new algorithm)
		pregnancy outcome (preterm, term) OLD ALGORITHM	C	tracking	delivdat, preterm, delvcode	c_case1 (old)	0=term 1=preterm .y=lost to follow-up
		pregnancy outcome (preterm, term) NEW ALGORITHM	C	tracking	delivdat, preterm, delvcode	c_case2 (new)	0=term 1=preterm .y=lost to follow-up
		randomly selected as a control	C	tracking	folweek	c_contrl	0=not control 1=control

* H=hierarchical, C=constructed

V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

		type*	source	source var name	constructed var name	constructed variable values
	composite case/control assessment OLD ALGORITHM	C	tracking		c_cascon (old)	0=cohort (not case or control) 1=case 10=control 11=control that became a case
	composite case/control assessment NEW ALGORITHM	C	tracking		c_cascn2 (new)	0=cohort (not case or control) 1=case 10=control 11=control that became a case
	birth weight	H	tracking	birthwt	c_bwt	in grams
	baby's race, based on parent race (if parents of diff. race, uses non-white; if parents of diff. race both non-white, uses mother's race; if father's race is missing, uses mother's)	C	main Q tracking	c_mrace, mqi7	c_brace	1=white 2=African American 3=American Indian 4=Asian or Pacific Islander <i>see attached list for other codes</i>
	LBW status	C	tracking	delvcode, birthwt	c_lbw	1= <1500 g 2= 1500 to <2500 g 3= 2500+ g .y=lost to follow-up
Small-for-gestational age and Large-for gestational age assessment All SGA and LGA assessments are limited to white and black babies, require parity, baby's gender, birthweight and gestational age to determine. Source: <i>Zhang J, Bowes Jr. WA. Birth-Weight-for-Gestational-Age Patterns by Race, Sex, and Parity in the United States Population. Obstet Gynecol 1995;86:200-8.</i>						
			tracking			

* H=hierarchical, C=constructed

V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

		type*	source	source var name	constructed var name	constructed variable values
SGA status OLD ALGORITHM		C		c_brace, c_bwt c_delges, c_liveb, gender	c_sga (old)	0=not SGA 1=SGA .x=all sources not available; cannot determine
SGA status NEW ALGORITHM		C	tracking	c_brace, c_bwt c_delgs2, c_liveb, gender	c_sga2 (new)	0=not SGA 1=SGA .x=all sources not available; cannot determine
LGA status OLD ALGORITHM		C	tracking	c_brace, c_bwt c_delges, c_liveb, gender	c_lga (old)	0=not LGA 1=LGA .x=all sources not available; cannot determine
LGA status NEW ALGORITHM		C	tracking	c_brace, c_bwt c_delgs2, c_liveb, gender	c_lga2 (new)	0=not LGA 1=LGA .x=all sources not available; cannot determine
Preterm subtype determination						
preterm subtype based on clinician determination OLD ALGORITHM			clinician determination		whyptb	0=term 1=preterm labor 2=preterm PROM 3=medically indicated 4=stillbirth 97=we have delivery date & it is preterm, but birth occurred at another site so determination not done 99=delivery code=3, no delivery info and well past due date
preterm subtype based on clinician determination NEW ALGORITHM					whyptb2	
clinician who made preterm subtype					ptbwho	1=Dr. John Thorp

* H=hierarchical, C=constructed

V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

	type*	source	source var name	constructed var name	constructed variable values
determination					2=Dr. Thad McDonald
clinician's confidence in the preterm subtype determination				ptbconf	1=very certain 2=less certain
Mom's weight gain during pregnancy (merged into the studyvar.x__ dataset; stored in \prom\data\sas\study\wtgain.x_) File also contains expected weight gain variable and gestational age at last weight variable					
Source of pre-pregnancy weight			screeener med abs	wtsource	1=screeener 2=medical abstract 3=imputed
weight gain total		med abs studyvar	wf02, c_lbs, wf07 uf02, uf11	c_wtgain	weight in kilograms
adequacy of total weight gain		med abs studyvar		c_IOMwt	ratio of observed weight over expected weight based on IOM guidelines (Institute of Medicine. Nutrition during pregnancy. Washington: National Academy Press, 1990.
Hypertension (merged into the studyvar.x__ dataset; stored in \prom\data\sas\study\hyper.x_)					
chronic hypertension		med abs	UG38, UF09a,UF09b WH02a,WF08a,WF08b	c_hyper	0 = not listed on chart 1 = listed on chart 9 = not in chart, insufficient test data . = no chart data or test results
pregnancy-induced hypertension		med abs	UG39, UA18 WE18	c_pih	
pre-eclampsia/eclampsia		med abs	UG40, UG41, UG42, UI04 WB03J, WB03K, WB03L, WB03T, WA09O, WA09P, WA09Q, WA09R	c_eclamp	
Diabetes					

* H=hierarchical, C=constructed

V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

	type*	source	source var name	constructed var name	constructed variable values
<i>(merged into the studyvar.x__ dataset, stored in \prom\data\sas\study\diabetes.x__)</i>					
This file also includes flag variables that identify the source of gestational diabetes assessment and the status of glucose tolerance test results					
gestational diabetes	C	med abs lab tests	we21, wl05, wl06, wl07, wl08, wl09, ua21, laboratory test results from clinic databases	c_gdm	0 = no 1 = yes 9 = not in chart, insufficient test data . = no chart data or test results
pre-existing diabetes	C	med abs	we20, wh07a ua20, ud11	c_pre_dm	
Anemia & Hematocrit					
Anemia in first trimester		med abs		c_anem1	0=not anemic 1=anemic
Anemia in second trimester		med abs		c_anem2	
Anemia in third trimester		med abs		c_anem3	
Elevated hematocrit in 1 st trimester		med abs		c_high1	0=hematocrit not high 1=high hematocrit
Elevated hematocrit in 2 nd trimester		med abs		c_high2	
Elevated hematocrit in 3 rd trimester		med abs		c_high3	
BV Assessment					
BV slide assay score, continuous				c_sb scor	
BV slide assay score, categorical				c_bvcat	1=negative, normal (score 0-3) 2=intermediate (score 4-6) 3=positive BV (score 7-10)
Self-reported reproductive tract & other infections during index pregnancy (constructed from multiple obs per woman file structure)					
yeast infection	C	main Q	mqinf typ, mqinf ted	c_mqb21	0-3 times
genital or vaginal warts	C	main Q	mqinf typ, mqinf ted	c_mqb27	0-3 times
genital herpes	C	main Q	mqinf typ, mqinf ted	c_mqb33	0-3 times

* H=hierarchical, C=constructed
V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

		type*	source	source var name	constructed var name	constructed variable values
	chlamydia	C	main Q	mqinftyp, mqinfted	c_mqb39	0-3 times
	trichomonas	C	main Q	mqinftyp, mqinfted	c_mqb45	0-3 times
	gonorrhea	C	main Q	mqinftyp, mqinfted	c_mqb51	0-3 times
	syphilis	C	main Q	mqinftyp, mqinfted	c_mqb57	0-3 times
	bacterial vaginosis	C	main Q	mqinftyp, mqinfted	c_mqb63	0-3 times
	pelvic inflammatory disease	C	main Q	mqinftyp, mqinfted	c_mqb69	0-3 times
Reproductive history (constructed from multiple obs per woman file structure in Main Q for PIN 1, dropped in PIN 2)						
	# previous pregnancies	H, C	main Q tracking	mqf1, mqf2 prevpreg	c_preprg	counts of reproductive events; 0 = never occurred or nulliparous
	# live births	C	main Q	mqf4	c_liveb	
	# miscarriages	C	main Q	mqf4	c_premis	
	# induced abortions	H, C	main Q tracking	mqf4 previndu	c_preabo	
	# previous stillbirths	C	main Q	mqf4	c_prestb	
	# previous known preterm births	H, C	main Q tracking	mqf6 prevptrm	c_preptb	
	# previous known LBW	C	main Q	mqf9a, mqf9b	c_prelbw	
	# previous ectopic	C	main Q	mqf4	c_precto	

* H=hierarchical, C=constructed
V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

		type*	source	source var name	constructed var name	constructed variable values
	# previous preterm PROM	C	main Q	mcf8	c_pprom	
	parity (live births + stillbirths)	H, C	main Q tracking	c_liveb + c_prestb prevterm + prevptm	c_parity	
	flag: # of pregs with imputed or recoded gestational age	C	main Q		c_gesflg	0-6 . if no main Q
	inter-pregnancy interval (between last and index)	C	main Q	pregnancy history section	c_prgint	months known implausible values of -2, -1, and 0; can be recoded to missing while waiting for med abs to correct
	flag: month of last preg was imputed	C	main Q		c_intflg	1=month was imputed
Behaviors during pregnancy						
	alcohol use at any time during pregnancy		main Q FUSP Q FUAD Q	mqc23 fpc14 fdc14		get from the Questionnaire source
	alcohol, 5 or more drinks/week after she knew she was pregnant		main Q FUSP Q FUAD Q	mqc24 fpc15 fdc15		get from the Questionnaire source
	other drugs during this pregnancy		main Q FUSP Q FUAD Q	mqc58 fpc29 fdc29		get from the Questionnaire source
	other drugs, any time in life		main Q	mqc56		get from the Questionnaire source

* H=hierarchical, C=constructed

V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

	type*	source	source var name	constructed var name	constructed variable values
average number of cigarettes smoked in months 1-6, continuous		main Q FUSP Q FUAD Q		c_cig16	continuous
average number of cigarettes smoked in months 1-6, dichotomous		main Q FUSP Q FUAD Q		c_cig16a	1=none, nonsmoker 2=yes, smoker
average number of cigarettes smoked in months 1-6, 4-levels		main Q FUSP Q FUAD Q		c_cig16b	1=1-9 cigarettes/day 2=10-19 cigs/day 3=20+ cigs/day 4=0, nonsmoker (high value reference for genmod)

Adequacy of Prenatal Care

<p>Kotelchuck's adequacy of prenatal care <i>(Kotelchuck M. An evaluation of the Kessner Adequacy of Prenatal Care Index and a proposed Adequacy of Prenatal Care Utilization Index. Am J Public Health 1994;84:1414-20.)</i> File also contains other variables pertaining to number of visits, expected visits, expected visit ratio, gestational age and month prenatal visits began.</p>					
Month prenatal care initiation index		med abs main q tracking studyvar		moindex4	Codes for all index variables: 0 = missing information 1 = inadequate 2 = intermediate 3 = adequate 4 = adequate plus
Expected visit index (categorical form of evratio)				evindex	
Summary of the two previous factors				indexsum	
Expected PNC visit ratio (observed/expected)				evratio	

Source file info for study variables

* H=hierarchical, C=constructed
 V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

		type*	source	source var name	constructed var name	constructed variable values
	Tracking file	C			c_trk	0=no, 1=yes
	Main Q files	C			c_mainq	0=no, 10=yes
	Main Q version date (to track why vars are missing)				mqverdat	
	Screeener file	C			c_screen	0=no, 100=yes
	Combination of the above source files	C			c_wehave	add the above for composite of source files available for observation (e.g., value 101 means we have tracking & screener, no Main Q)
Study administrative variables						
	Site of recruitment – UNC or Wake				c_site	1=UNC 2=Wake
	Cohort determination – PIN 1 or PIN 2				c_pin1_2	1=PIN 1 2=PIN 2
	Multiple pregnancies in the PIN study	C	tracking		c_pindup	need to update manually in the program; first digit is the # of the preg in PIN, other digits are woman link: e.g., 1001 & 2001 are the same woman, pregs 1&2; 1012, 2012, 3012 are same woman, pregs 1, 2, & 3
	Medical abstract – different abstract efforts were done for various vars; methodologies for assessing varied				co_flag	

* H=hierarchical, C=constructed

V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

	type*	source	source var name	constructed var name	constructed variable values
Indicates that Consent for genetic testing was provided on the PIN1/2 Consent form. Note: Also see the C_WBC variable	C	PIN1/2 Consent Form		c_dna	0=NO 1=YES
Indicates that a blood biospecimen, white blood cells, suitable for genetic (DNA) testing was collected and is available as of 4/1/2009. Note: Also see the C_DNA variable	C	PIN1/2 Biospecimen Inventory		c_wbc	0=NO 1=YES
Physical activity variables					
Any activity (excluding other) - 3 months before pregnancy	C	main q	mqj2a, mqj3a, mqj4a mqj5a,mqj6a	patot1	0 = no activity 1 = some activity
Any activity (excluding other) - 1 st trimester	C	main q	mqj2a, mqj3a, mqj4a mqj5a,mqj6a	patot2	0 = no activity 1 = some activity
Any activity (excluding other) - 2 nd trimester	C	main q	mqj2a, mqj3a, mqj4a mqj5a,mqj6a	patot3	0 = no activity 1 = some activity
Total activity (excluding other) - 3 months before pregnancy	C	main q	mqj2b, mqj3b, mqj4b mqj5b, mqj6b	patot1t	hours/week
Total activity (excluding other) - 1 st trimester	C	main q	mqj2b, mqj3b, mqj4b mqj5b, mqj6b	patot2t	hours/week
Total activity (excluding other) - 2 nd trimester	C	main q	mqj2b, mqj3b, mqj4b mqj5b, mqj6b	patot3t	hours/week
Any activity (including other) - 3 months before pregnancy	C	main q	mqj2a, mqj3a, mqj4a mqj5a,mqj6a, mqj7a	patoto1	0 = no activity 1 = some activity
Any activity (including other) - 1 st trimester	C	main q	mqj2a, mqj3a, mqj4a mqj5a,mqj6a, mqj7a	patoto2	0 = no activity 1 = some activity

* H=hierarchical, C=constructed
V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

		type*	source	source var name	constructed var name	constructed variable values
	Any activity (including other) - 2 nd trimester	C	main q	mqj2a, mqj3a, mqj4a mqj5a, mqj6a, mqj7a	patoto3	0 = no activity 1 = some activity
	Total activity (including other) - 3 months before pregnancy	C	main q	mqj2b, mqj3b, mqj4b mqj5b, mqj6b, mqj7b	patoto1t	hours/week
	Total activity (including other) - 1 st trimester	C	main q	mqj2b, mqj3b, mqj4b mqj5b, mqj6b, mqj7b	patoto2t	hours/week
	Total activity (including other) - 2 nd trimester	C	main q	mqj2b, mqj3b, mqj4b mqj5b, mqj6b, mqj7b	patoto3t	hours/week

* H=hierarchical, C=constructed

V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

The following are date variables that are needed to construct various intervals, but because of deductive disclosure, they are NOT distributed for analysis.

Mom's date of birth			mqq1a, mqq1b, mqq1c; momage24	c_momdob	SAS date format
Date of delivery				delivdat	
Date recruited into PIN				recruitd	SAS date format
LMP equivalent start of pregnancy using new dating algorithm (startprg used old algorithm and is in the tracking files)				strtrpg2	
<i>TRIMESTER DATING USING OLD ALGORITHM</i>					
Beginning of 1 st trimester				c_1tri_b	
End of 1 st trimester				c_1tri_e	
Beginning of 2 nd trimester				c_2tri_b	
End of 2 nd trimester				c_2tri_e	
Beginning of 3 rd trimester				c_3tri_b	
End of 3 rd trimester				c_3tri_e	
<i>TRIMESTER DATING USING NEW ALGORITHM</i>					
Beginning of 1 st trimester using strtrpg2				c_1trib2	
End of 1 st trimester using strtrpg2				c_1trie2	
Beginning of 2 nd trimester using strtrpg2				c_2trib2	
End of 2 nd trimester using strtrpg2				c_2trie2	

* H=hierarchical, C=constructed
V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX

	Beginning of 3 rd trimester using strtprg2				c_3trib2	
	End of 3 rd trimester using strtprg2				c_3trie2	

RACE variables: c-mrace, c_brace

01	white
02	African American (black or African)
03	American Indian or Native American
04	Asian or Pacific Islander
05	other (not specified)
06	White Hispanic
07	Black Hispanic
08	Hispanic
09	White/Black (or Black/White)
10	Asian/Black (or Black/Asian)
11	Indian/Cuban
12	White/Native American
13	Black/Native American
14	Black/Hispanic/Jamaican
15	Asian/white or Asian American

* H=hierarchical, C=constructed
V:\PROMDATA\DOCUMENT\STUDY\STUDYVAR.DOCX

16	Canadian
17	African American/Canadian
18	Portuguese or Brazilian
19	

[see \prom\data\document\codebook.pin]

* H=hierarchical, C=constructed
V:\PROMDATA\DOCUMENT\STUDY\STUDYVAR.DOCX

Tips to Analysts

1. Check the subgroup classification based on c_cascon or c_sga (etc.)
DO NOT depend on any master file having a limited group (e.g., the medical abstracts have more than preterm cases and controls)
2. We have women with more than one pregnancy in the study – see c_pindup
– these observations are not independent
3. When available, use study-wide variables

* H=hierarchical, C=constructed
V:\PROM\DATA\DOCUMENT\STUDY\STUDYVAR.DOCX