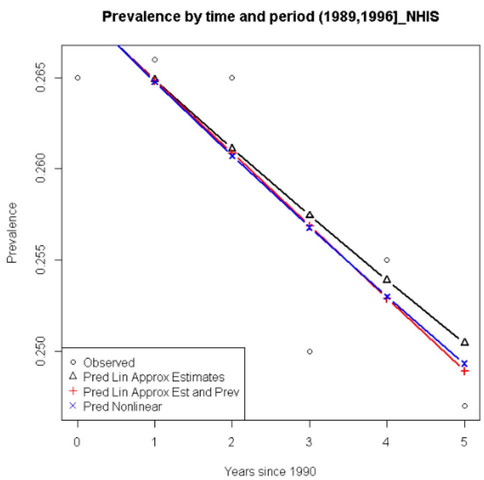
Notes on the analysis.

The estimation process was performed using **r** statistical software. First, the intercept term pi(0) needs to be estimated as an unknown parameter together with theta. Second, plain nonlinear regression is not optimal (does not provide efficient estimates that deliver smallest possible standard errors and confidence intervals). What we did is a regression where squared differences between observed and predicted values were weighted by the inverse of the variance of prevalence estimates (squared standard errors) that come from Table 1. We also performed the estimation separately for each period and data source. Estimated parameters are presented in the following table.

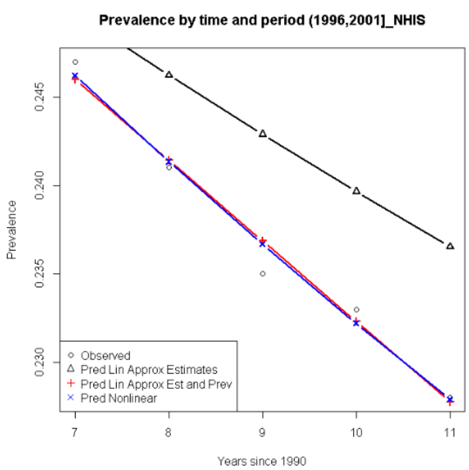
Parameter	Estimate	StdError	Statistic	pValue	period_data
pi0	0.268948	0.002949	91.2005	8.67E-08	(1990,1995]_NHIS
theta	0.023913	0.004649	5.144142	0.006772	(1990,1995]_NHIS
pi0	0.286779	0.004326	66.29612	7.56E-06	(1997,2001]_NHIS
theta	0.033742	0.001491	22.63122	0.000189	(1997,2001]_NHIS
pi0	0.30062	0.018776	16.01098	8.90E-05	(2002,2007]_NHIS
theta	0.032676	0.00395	8.27236	0.001165	(2002,2007]_NHIS
pi0	0.442708	0.032178	13.75813	3.64E-05	(2008,2014]_NHIS
theta	0.044571	0.003074	14.49734	2.82E-05	(2008,2014]_NHIS
pi0	0.317812	0.016917	18.78608	4.73E-05	(2002,2007]_NSDUH
theta	0.032366	0.002999	10.79362	0.000418	(2002,2007]_NSDUH
pi0	0.43625	0.023675	18.42649	8.66E-06	(2008,2014]_NSDUH
theta	0.04183	0.002049	20.41215	5.22E-06	(2008,2014]_NSDUH

Table uses raw scale; multiply by 100 to get it in % scale.

We originally intended to apply a non-linear meta-regression throughout to test for an increase in the cessation rate, but could not identify software that would allow us to do this for the custom model at hand. We could do it though if the model were linear. Having experimented with the linearization of the model we found that while the linearized model does fit well, its agreement with the original model is not good for some periods. To check that we obtained estimates of pi0 and theta for the linearized model and plugged them into the linearized model. This model/method always fits well and similar to the nonlinear procedure that has both estimates and predictions done on the non-linear original model. Plugging estimates obtained using linearized model into the original model does not always fit well. The story is summarized in the following plots.



This is an example where all methods fit well.



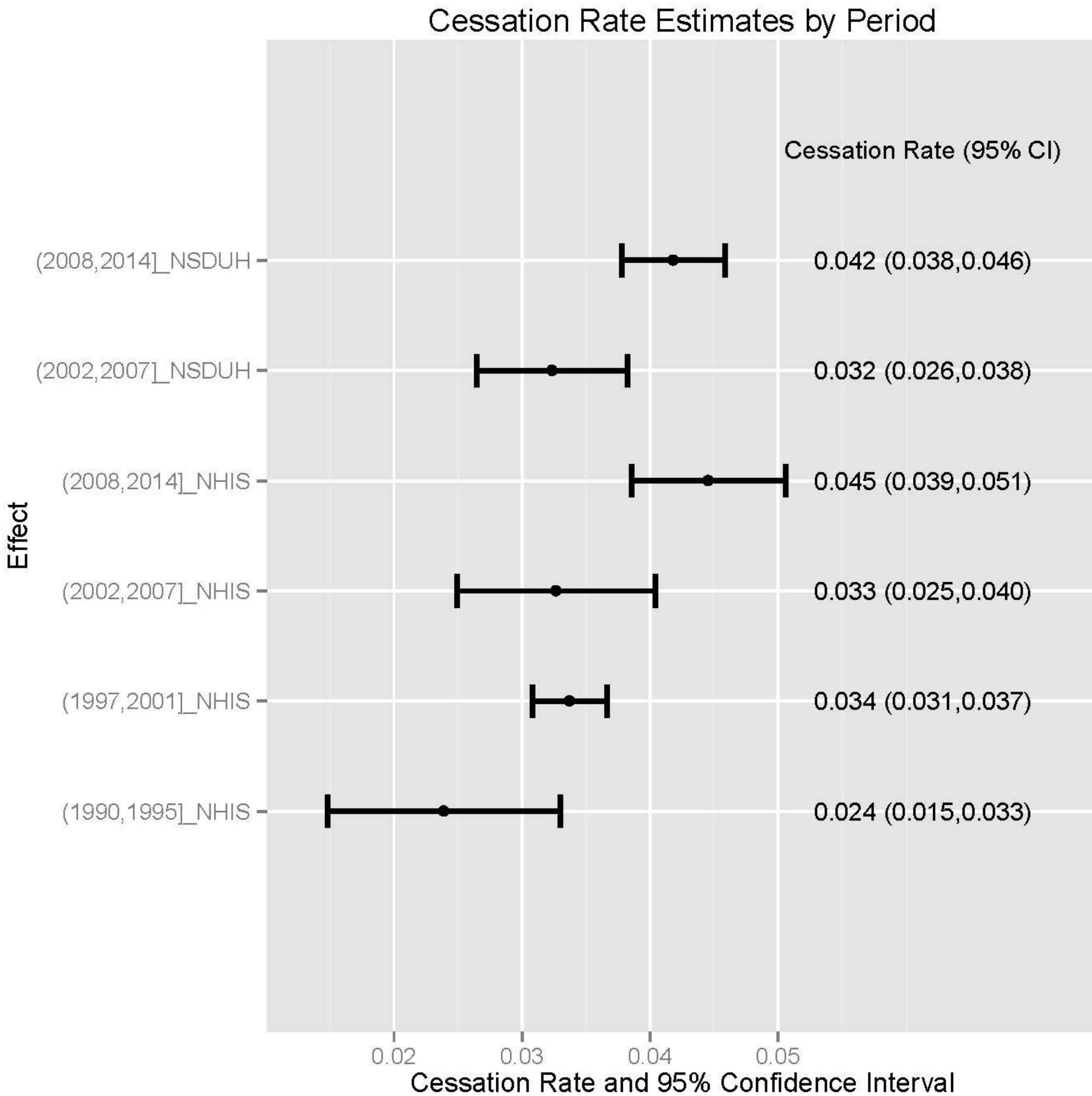
This is an example where linearized estimates do not go well with the original model (upper line).

In view of the above analysis we decided to go in two stages:

1. Get correct and efficient estimates of theta and pi0 using inversely weighted nonlinear least squares regression for each period and data source combination separately (done above).
2. Do a linear meta-regression across periods and data sources on estimates of theta as a response (with their predicted standard errors), and year since 1990 as continuous covariate, and data source as a categorical covariate.

Meta-Regression.

The estimates of cessation rates by time and data source obtained from the nonlinear weighted regression are given in the following figure.



The results of meta regression are shown in the following table

Parameter	Estimate	StdError	Statistic	pValue	95%ci.lb	95%ci.ub
(Intercept)	0.026613	0.002216	12.01066	0	0.02227	0.030956
t	0.005998	0.001373	4.367799	1.26E-05	0.003306	0.008689
dataNSDUH	-0.00388	0.002781	-1.39478	0.163082	-0.00933	0.001572

Note that there is a **highly significant trend in cessation rate** (p=1.26E-05). **The effect of data source is nonsignificant** (p = 0.16). The model fits well as shown on the residual plot and the observed vs predicted plot.