

Validating risk factor and chronic disease projections in the Future Adult Model

Bryan Tysinger^{1*}

¹University of Southern California, Los Angeles, California

Abstract Over the past several decades, the United States has experienced a dramatic rise in obesity rates, due to both a rightward shift of the body mass index (BMI) distribution and a pushing out of the right tail. This shift has led to increases in obesity-related chronic diseases, particularly diabetes, as well as impacts on longevity, medical expenditures, and quality of life. Microsimulation modeling is a potentially useful tool for assessing the impacts of policies targeting this epidemic, but reliably assessing policies requires a model that performs well in projecting health risk factors and disease outcomes. This research assesses the out-of-sample and external validity of a microsimulation model of the U.S. adult population. There are two research questions addressed in this analysis: 1. How well does the Future Adult Model (FAM) perform in projecting BMI and diabetes over a ten-year horizon compared to the host data? 2. How well do the microsimulation model's predictions compare to external surveillance data of BMI and diabetes? FAM is an economic-demographic microsimulation model of the United States population over the age of 25. For this validation exercise, all Markov transition models are estimated using the 1999-2007 waves of the PSID. The simulation is then run from 2007-2017. For internal consistency, simulated outcomes in 2017 are compared to actual PSID outcomes. Population means and selected quantiles are compared between the simulation and the host data. Receiver operating characteristic (ROC) curves are used to assess model performance for binary outcomes using the area under the curve (AUC) statistic. For external validation, simulated outcomes for 2007-2017 are compared to the Behavioral Risk Factors Surveillance System (BRFSS), a large, nationally-representative survey of the United States population. After ten years of simulation, FAM BMI projections for men and women compare well to both PSID and BRFSS data throughout much of the distribution. The 99th percentile differs significantly, with FAM underestimating the right tail of the BMI distribution. Individual assignment of obesity and severe obesity performs well using AUC as a criteria. Initial differences in the diabetes prevalence between PSID and BRFSS data are preserved in FAM projections. FAM is initially 1.9 percentage points below BRFSS for women 25 and older and is 1.6 percentage points below BRFSS for women 35 and older after ten years of simulation. Men 25 and older are 1.2 percentage points lower initially and are 0.8 percentage points lower after ten years of simulation. Individual assignment of diabetes incidence does not perform as well as clinical models with richer predictors. Researchers using FAM should be cognizant of these strengths and limitations of the microsimulation model.

*For correspondence:
btylinge@usc.edu

©This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Author Keywords:
Microsimulation, health economics, life-course model, validation
© 2020, Bryan Tysinger.

JEL classification: C6, I1, J1

DOI: <https://doi.org/10.34196/ijm.00225>

1. Introduction

Over the past several decades obesity rates (a body mass index (BMI) over 30) in the United States have increased dramatically, from 22.9% in 1988-1994 to 39.6% in 2015-2016 (*Fryar et al., 2019*). During the same time period, the severely obese prevalence (BMI over 40) increased from 2.8% to 7.7%. Comorbidities associated with obesity, such as diabetes, have also increased (*Menke et al., 2015*). These increases impact both individuals and society through many pathways, including the direct health impacts (chronic

conditions, functional limitations, and mortality), increased medical expenditures and utilization, other economic impacts (changes in employment, disability), and declines in subjective well-being.

Many strategies have been suggested to tackle this challenge, including lifestyle interventions that target diet and exercise, medical procedures like gastric bypass surgery, pharmaceuticals that lead to weight loss, taxes on particular foods, menu labeling to improve food choices, and more. These strategies, in turn, are often implemented on a small scale or within a randomized control trial. To translate the effects to a larger scale analysts increasingly turn to microsimulation models, as these models can account for heterogeneous individuals and treatment effects. To do this credibly requires a microsimulation model that performs well, but also one where the limitations are clearly conveyed. With respect to body mass index, a model that performs well needs to project not just the mean BMI well, but also capture the distribution, as policies often target those in the right tail, such as the obese or severely obese. This is not an uncommon requirement for microsimulation models, as the distribution of continuous variables is often relevant to policy questions. Though often assumed to be “black boxes” or “crystal balls” for projecting the future, validation of past-performance and transparency about limitations can avoid overselling the capabilities of these models while better informing policy decisions.

There are two purposes for this paper. The first is to describe an approach to validation used with the Future Adult Model (FAM), a microsimulation model of the United States population over age 25. This approach is potentially applicable to other microsimulation models with similar data availability. The second goal is the validation itself, highlighting both where the model performs well and areas where further research can improve FAM’s projections.

Section 2 briefly describes the model and its source data. A few approaches to out-of-sample validation are presented in Section 3. External validation is described in Section 4. Section 5 concludes.

2. Data and methods

FAM is described in a detailed technical appendix (<https://healthpolicy.box.com/v/FAM-appendix-2018>), so only the core functions of the model are summarized here. The host data for the simulation

Table 1. Variables directly impacted by BMI and diabetes in FAM.

Domain	Category	Measure
Health	Chronic conditions	Cancer ¹ , diabetes ¹ , heart disease, hypertension, lung disease ¹ , stroke
	Functional limitations	Activities of daily living, instrumental activities of daily living
	Mental distress	Kessler 6
	Mortality	Death ²
	Risk factors	BMI ¹ , start smoking, stop smoking
Economic	Employment status	Full-/part-time, labor force participation
	Health insurance	Health insurance type ²
	Income and assets	Capital income ² , earnings ² , wealth
	Public program participation	OASI ² , DI ²
Medical cost and use	Individual	Drug \$ ² , out of pocket \$ ²
	Medicaid	\$ ²
	Medicare	Total \$ ² , Part A \$ ² , Part B \$ ²
	Total expenditures	\$ ²
Subjective well-being	Utilization	Doctors visits ² , hospital encounters ² , hospital nights ²
		Life satisfaction ² , quality-adjusted life years ² , self-reported health ²

1 BMI model only

2 Diabetes model only

is the Panel Study of Income Dynamics (PSID), a biennial panel survey which collects information on a broad set of health and economic outcomes for approximately 15,000 respondents per year (McGonagle et al., 2012). While the PSID is designed to be nationally representative, it does face the challenges all longitudinal surveys do, such as imperfect response rate and sample attrition. We typically use PSID data from 1999-2017 as the basis for FAM. These data are used to estimate transition functions and as the initial data for the simulation. The first-order Markov transition models are the engine used in “aging” the simulants. The transitions are a mixture of continuous, binary, and categorical outcomes, with a time-scale that mimics the two-year structure of the data. FAM simulates dozens of variables for individuals, including health risk factors, chronic conditions, functional limitations, mortality, life events, economic outcomes, medical cost and use, and government transfers. The causal pathway of FAM is that health risk factors (such as body-mass index and smoking) impact health outcomes (such as diabetes incidence or the number of functional limitations), which in turn impact a set of economic outcomes (such as medical expenditures, retirement).

BMI and diabetes are two critical outcomes, but also enter as predictors for many other models. These pathways are summarized in **Table 1**. Since both BMI and diabetes impact so many things, it is critical to understand the quality of projections as errors will propagate. The distribution of BMI matters for these outcomes, as high BMI increases risk of chronic illness, so ideally the simulation will capture the distribution of BMI, not just the mean.

Note that both of these measures are reported by survey respondents, not based on clinical measurements or administrative records. Self-report data are known to suffer from biases in reporting. Individuals often under-report weight and over-report height, leading to BMI values that are too low compared to their measured values. Similar challenges face measures of chronic conditions, though chronic conditions are more challenging to assess in an interview setting. Administrative medical claims are not a panacea, as they have their own challenges (Clair et al., 2017). With that said, the measures used in FAM are commonly used in surveys such as PSID, the Behavioral Risk Factor Surveillance System (BRFSS), the Medical Expenditure Panel Survey (MEPS), the Medicare Current Beneficiary Survey (MCBS), the National Health Interview Survey, and others. Though they do not provide the true population prevalence of chronic conditions or BMI distribution, they are still useful in many ways. Within FAM, medical expenditures are estimated using MEPS and MCBS, which then allows the translation of these self-reported measures into predicted medical expenditures.

FAM estimates the two-year transition of BMI. BMI is defined as an individual’s mass divided by the square of their height. Clinically, it is often used as a predictor of subsequent clinical outcomes, such as risk of diabetes or mortality. The transition model for BMI is estimated in natural logs, allowing the interpretation as a percent change in BMI. A Box-Cox analysis of BMI transitions in the PSID data suggests a λ parameter between 0 and -0.5 , depending on specification. The transition model is of the form:

$$\ln(BMI_{i,t+1}) = \alpha \cdot \ln(BMI_{i,t}) + X_{i,t} \cdot \beta + Y_i \cdot \gamma + \epsilon_{i,t+1} \quad (1)$$

The transition model estimates for BMI, estimated separately for men and women, are shown in **Table 2**. These are reduced-form models that include both time-varying (age, log BMI with several knots, marital status) and static (race/ethnicity, education, characteristics from early age) predictors. Time-varying predictors enter as two-year lagged variables. One can think of the interpretation of this model as a percent change in BMI over a two-year period. Within the simulation, a random draw from the root-mean square error term of this model adds a stochastic element. BMI appears in other transitions either in logs or in clinically-relevant BMI categories (under 25, 25 to 30, 30 to 35, 35 to 40, or over 40).

FAM’s diabetes model is a probit model of two-year diabetes incidence. As an incidence model, only individuals who did not have diabetes in the previous period are included in the estimation.

$$Pr(Diabetes_{s,t+1} = 1|X) = \phi(X_{i,t} \cdot \beta + Y_i \cdot \gamma + \epsilon_{i,t+1}), \text{ if } Diabetes_t = 0 \quad (2)$$

Here, we adjust for time-varying covariates (age, smoking status, exercise, and the log of the previous BMI in splines with several knots) and static characteristics (race/ethnicity, education, gender, childhood SES, childhood health). Marginal effects of this model are shown in **Table 3**. Consistent with the wording of the question in the PSID (“Has a doctor or health professional ever told you that

Table 2. Two-year BMI transition estimates used in FAM, 1999-2007 PSID.

	(1) Log(BMI)-Females	(2) Log(BMI)-Males
Non-Hispanic Black	0.00838***	0.00168
Hispanic	0.00403	0.000181
Less than HS	0.00421	-0.000211
Bachelors	-0.00887***	-0.00585***
Masters or higher	-0.00741*	-0.00460
Non-Hispanic Black× Less than HS	-0.00920*	-0.00877*
Non-Hispanic Black× Bachelors	0.0114*	0.00670
Non-Hispanic Black× Masters or higher	-0.00390	0.0210
Hispanic× Less than HS	-0.00298	-0.00431
Hispanic Bachelors	0.00193	-0.000211
Hispanic× Masters or higher	-0.00810	-0.00361
Poor child SES	0.00103	0.00121
Well-off child SES	-0.00154	-0.00146
Fair child health	0.00489	-0.000600
Good child health	0.000692	0.000749
Very good child health	0.00254	-0.000955
Excellent child health	0.000864	-0.00229
Age spline, less than 35	-0.000382	-0.0000346
Age spline, 35 to 44	-0.000155	-0.000609*
Age spline, 45 to 54	0.0000296	-0.000446
Age spline, 55 to 64	-0.000515	0.000165
Age spline, 65 to 74	-0.00130**	-0.00147***
Age spline, more than 75	-0.00202***	-0.00117*
Lag of Log(BMI) spline, BMI less than 20	0.769***	0.327***
Lag of Log(BMI) spline, BMI 20 to 25	0.945***	0.929***
Lag of Log(BMI) spline, BMI 25 to 30	0.898***	0.905***
Lag of Log(BMI) spline, BMI 30 to 35	0.987***	0.945***
Lag of Log(BMI) spline, BMI 35 to 40	0.805***	0.872***
Lag of Log(BMI) spline, BMI over 40	0.895***	0.830***
Cohabiting	-0.00124	0.00256
Married	-0.00480**	0.000316
Constant	0.738***	2.057***
Observations	20942	16454
R ²	0.836	0.810

***p < 0:05, *p < 0:01, **p < 0:001

you had diabetes or high blood sugar?”), diabetes is treated as an absorbing state variable for the remainder of a simulant’s life.

In typical FAM use, all recent waves of PSID data are pooled for estimation. However, for the analyses presented here we restrict the data used for estimation to PSID respondents from 1999 to 2007.

Table 3. Two-year diabetes incidence estimates used in FAM, 1999-2007 PSID.

	(1) Diabetes incidence (marginal effects)
Non-Hispanic Black	0.00167
Hispanic	0.00505
Less than HS	0.00331
Bachelors	-0.000224
Masters or higher	-0.00235
Male	0.00268*
Poor child SES	0.00322*
Well-off child SES	0.00150
Fair child health	-0.000465
Good child health	-0.00240
Very good child health	0.00184
Excellent child health	0.00175
Age spline, less than 35	0.0000936
Age spline, 35 to 44	0.00117***
Age spline, 45 to 54	0.000778**
Age spline, 55 to 64	0.000295
Age spline, 65 to 74	0.000269
Age spline, more than 75	-0.000564
Lag of former smoker	0.00141
Lag of current smoker	0.00550**
Lag of any exercise	-0.00394*
Lag of Log(BMI) spline, BMI less than 25	0.0485**
Lag of Log(BMI) spline, BMI 25 to 30	0.0733***
Lag of Log(BMI) spline, BMI 30 to 35	0.0382*
Lag of Log(BMI) spline, BMI 35 to 40	0.0666**
Lag of Log(BMI) spline, BMI over 40	-0.0235
Observations	35264
Pseudo R ²	0.117

***p < 0:05, *p < 0:01, **p < 0:001

Using these transition models, we simulate the 2007 PSID sample through 2017. The 2017 projections are then compared to the 2017 PSID to test consistency with the host data and to an external data source, the 2017 Behavioral Risk Factor Surveillance System (BRFSS), for external validity. Uncertainty in the transition models is handled via non-parametric bootstrapping. Fifty sets of transition models are estimated on resampled PSID data. Each set of transition models is then simulated 100 times. Confidence intervals reflect both transition and Monte Carlo uncertainty.

3. Out-of-sample validity

We explore FAM's out-of-sample validity in two ways. The first focuses on population-level statistics for the two outcomes of interest. The second uses Receiver Operating Characteristic (ROC) curves to assess how well FAM classifies individuals compared to their actual outcomes on a 10-year horizon.

Table 4. Out-of-sample validation - 2017 FAM vs. 2017 PSID.

	Females		Males	
	FAM (2017)	PSID (2017)	FAM (2017)	PSID (2017)
BMI				
1st pctl	16.6 [16.2, 16.9]	17.2 [16.7, 17.8]	19.2 [19.0, 19.5]	19.0 [18.5, 19.5]
5th pctl	18.7 [18.5, 19.0]	19.5 [19.2, 19.8]	21.2 [21.0, 21.4]	21.6 [21.3, 21.9]
10th pctl	20.1 [20.0, 20.3]	20.5 [20.4, 20.7]	22.5 [22.2, 22.7]	22.8 [22.5, 23.0]
25th pctl	22.9 [22.8, 23.0]	22.8 [22.5, 23.1]	24.9 [24.7, 25.0]	24.9 [24.7, 25.1]
Mean	27.7 [27.6, 27.9]	27.6 [27.4, 27.9]	28.5 [28.3, 28.7]	28.3 [28.1, 28.6]
75th pctl	31.5 [31.3, 31.8]	30.9 [30.5, 31.3]	31.5 [31.2, 31.8]	30.9 [30.5, 31.3]
90th pctl	36.7 [36.3, 37.0]	36.6 [35.9, 37.2]	35.3 [34.8, 35.7]	35.1 [34.6, 35.6]
95th pctl	40.0 [39.4, 40.5]	40.8 [40.1, 41.4]	37.7 [37.1, 38.2]	37.8 [37.3, 38.4]
99th pctl	46.9 [45.5, 48.2]	50.5 [48.1, 52.9]	42.4 [41.3, 43.6]	46.4 [44.6, 48.1]
Diabetes prevalence	12.4 [11.5, 13.3]	14.6 [13.2, 15.9]	14.9 [13.6, 16.3]	16.7 [15.2, 18.2]

Notes: Confidence intervals for FAM reflect 50 sets of bootstrapped transition models, each simulated 100 times. Confidence intervals for PSID reflect the complex survey design.

3.1. Aggregate projections

Beginning with the 2007 cohort of 25 and older PSID respondents, we simulate through 2017. Selected projections for the 2017 cohort, now 35 and older, are presented in **Table 4**.

For women, mean BMI is approximately the same between FAM projections (27.7) and PSID respondents (27.6) for 2017. For men, projected mean BMI is 28.5 in FAM, 28.3 for PSID. Most percentiles also compare well, with overlapping confidence intervals. In the right tail, such as at the 95th and 99th percentile, FAM and the PSID begin to deviate. At the 99th percentile, the FAM projection is 3.6 BMI points lower than the PSID value. For men, we see a similar story, though the discrepancy is around 4.2 BMI points at the 99th percentile. This suggests that the 1999-2007 data used for estimating the transition models did not accurately capture the continual expansion of the very high BMI population in the US.

Diabetes prevalence is lower in FAM than the PSID for both women and men. FAM projects 12.4% prevalence for women 35 and older, compared to the 14.6% observed in the PSID. Similarly, FAM projects 14.9% diabetes prevalence for men compared to 16.7% in the 2017 PSID. This is possibly a consequence of not projecting the right-tail of the BMI distribution. Alternatively, it could be changing diagnostic practices in the US between 2007 and 2017. A temporal time-trend would not be captured with FAM's approach to transition model estimation.

3.2. Individual assignment

In order to assess FAM's performance in classifying individuals, we use Receiver Operating Characteristic curves, as proposed in a validation of a cardiovascular disease microsimulation model (**Pandya et al., 2017**). In this analysis, each 2007 PSID respondent is simulated 5,000 times (50 sets of bootstrapped transition models, each with 100 Monte Carlo replications). For the outcome of interest, the fraction of simulations that resulted in that outcome is calculated. All simulants are ordered, smallest to largest, by these fractions. At each level of the fraction, the ROC analysis compares to the observed outcome for the individuals in the PSID survey data, assessing the true positive rate and false positive rate. The ROC analysis then shows the trade-off between "true positives" and "false negatives" for different thresholds of predicted prevalence after ten years. We assess four outcomes: having a BMI under 25, having a BMI over 30, having a BMI over 40, and 10-year incident diabetes for those who did not have diabetes in 2007. A performance parameter is the "area under the curve" (AUC in the figures). Interpretation of the value of the AUC statistic varies by context. An AUC of 1.0 is perfect classification, 0.5 is no better than random chance.

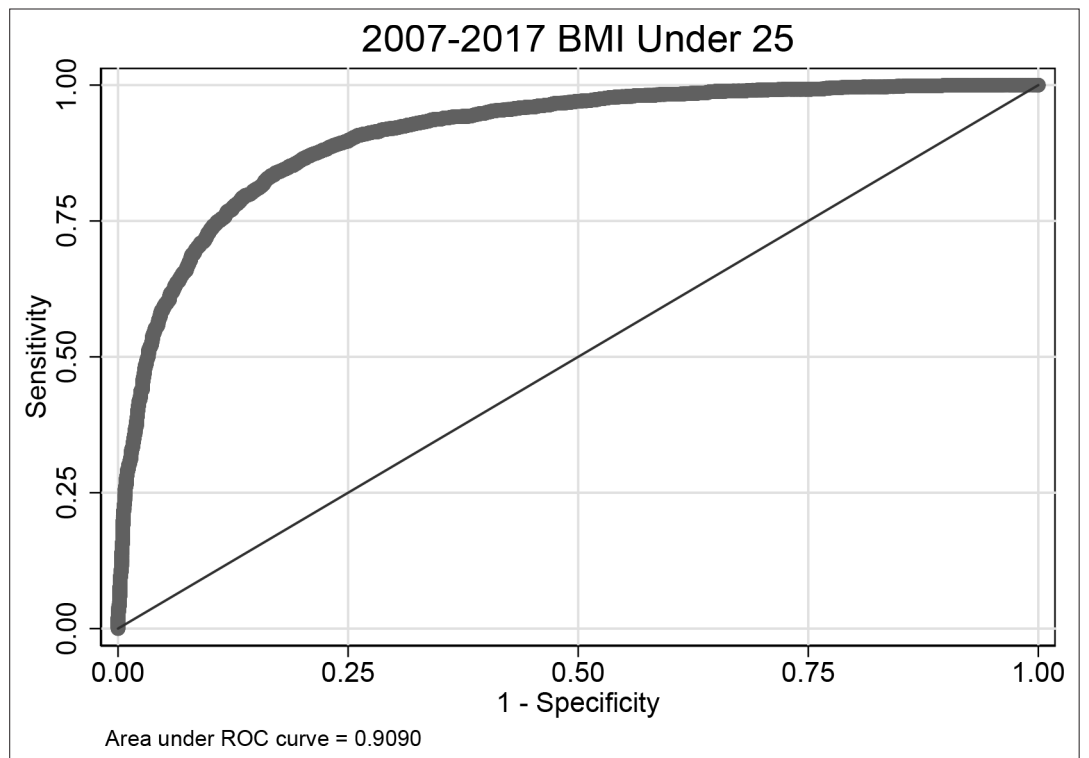


Figure 1. Receiver Operating Characteristic for BMI \leq 25 - Females

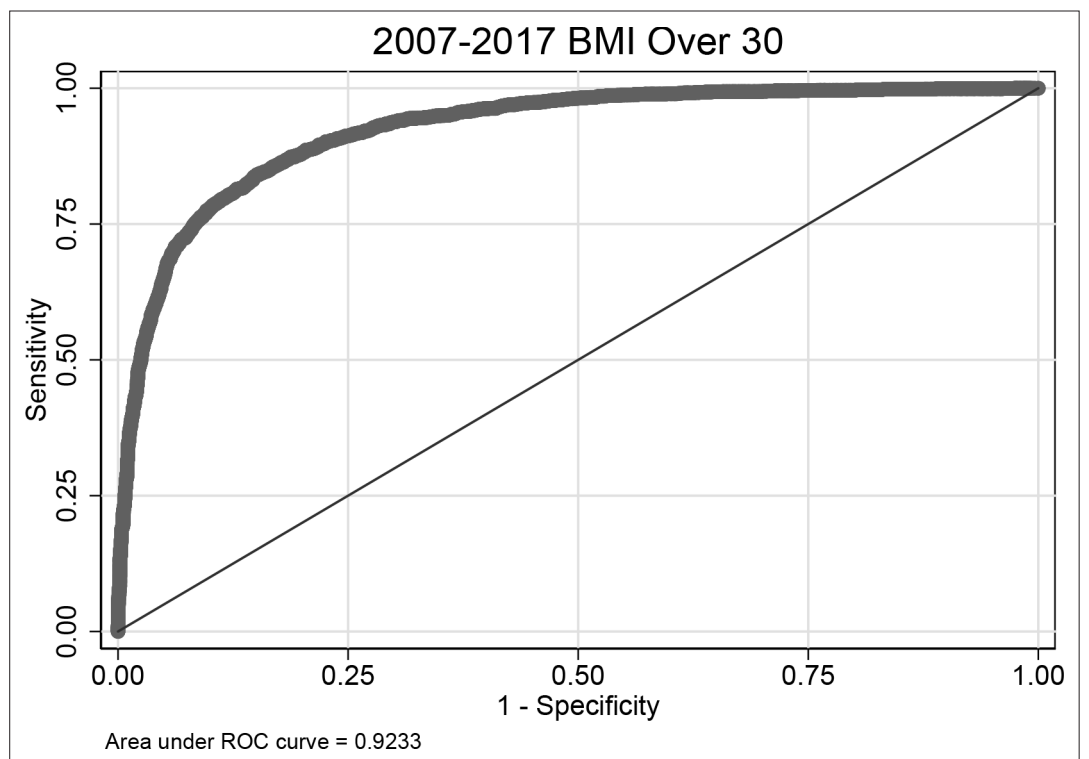


Figure 2. Receiver Operating Characteristic for BMI $>$ 30 - Females

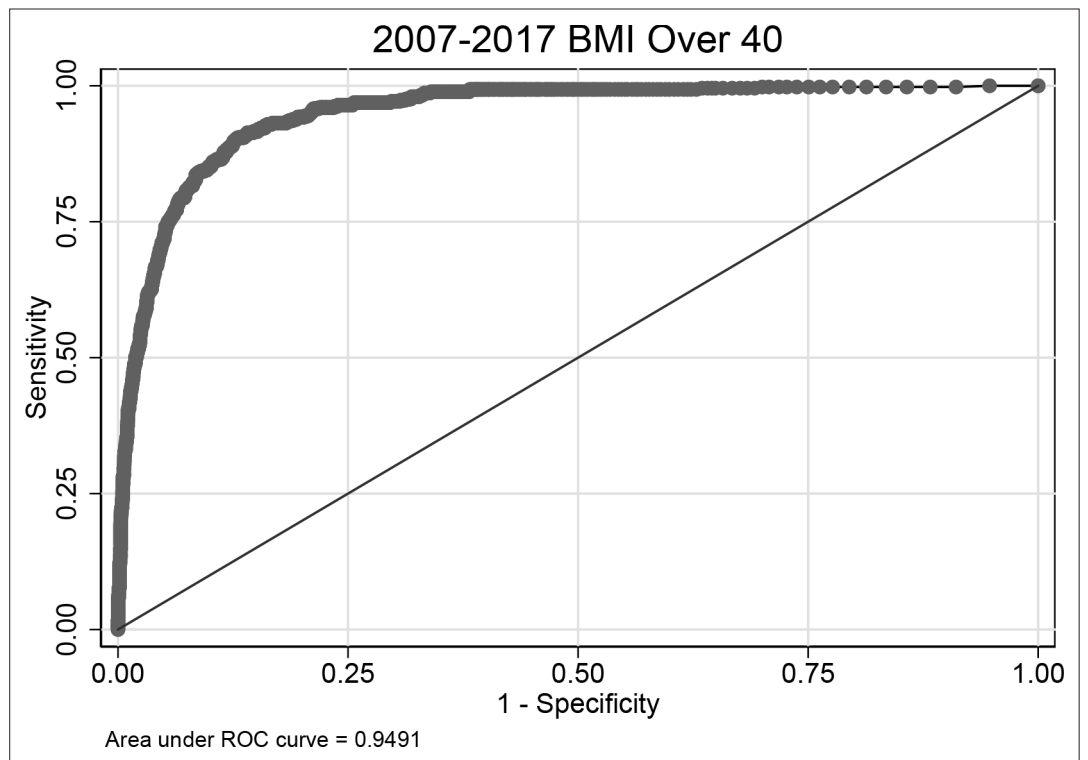


Figure 3. Receiver Operating Characteristic for BMI > 40 - Females

For females, the AUC for predicting a BMI under 25 is 0.91 (**Figure 1**), while the AUC for predicting a BMI over 30 is 0.92 (**Figure 2**). For predicting a BMI over 40, the AUC is 0.95 (**Figure 3**). These

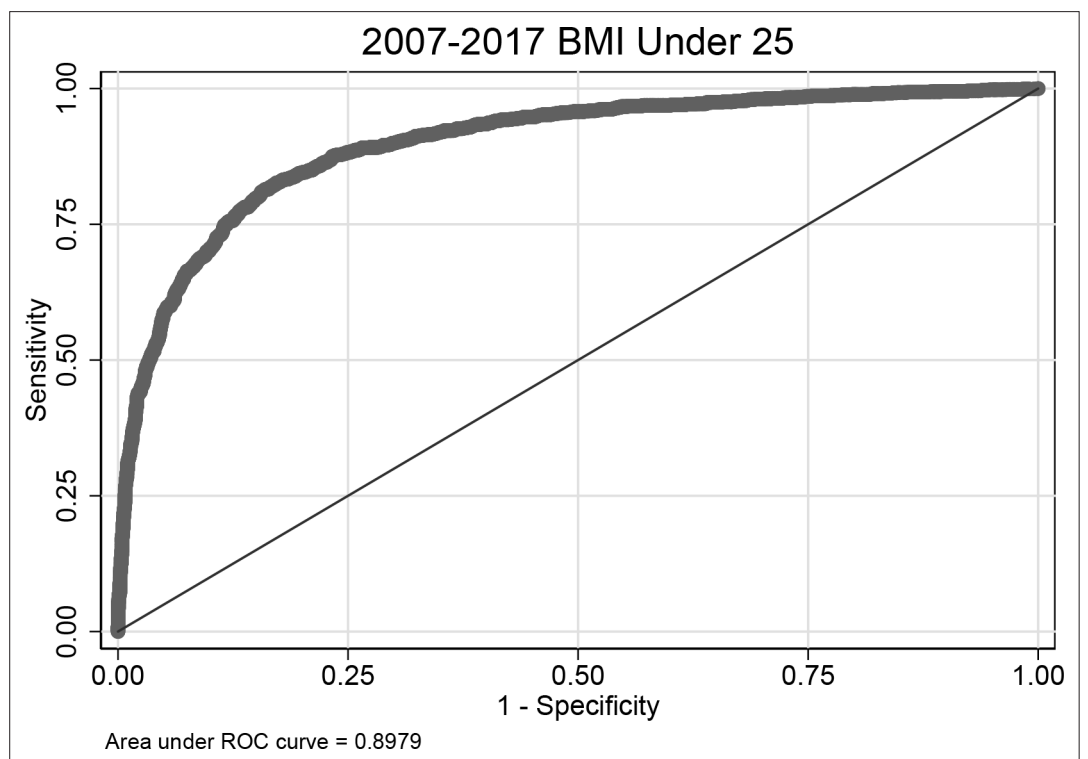


Figure 4. Receiver Operating Characteristic for BMI ≤ 25 - Males

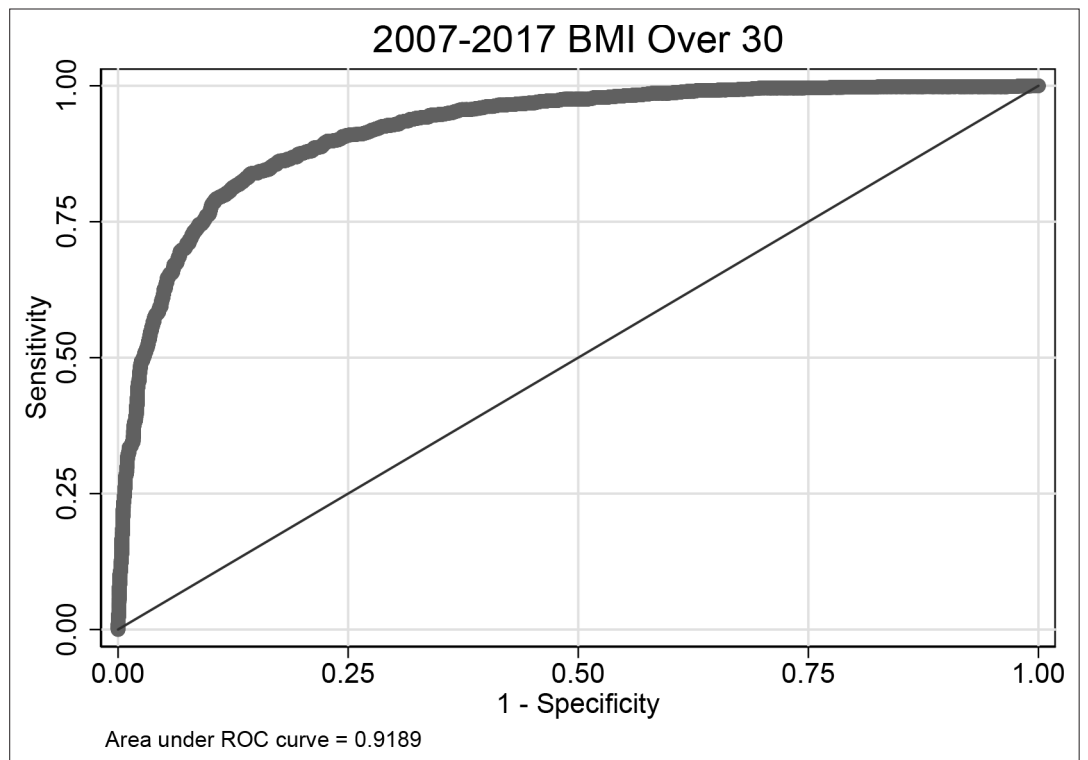


Figure 5. Receiver Operating Characteristic for BMI > 30 - Males

perform well, likely driven by the persistence of an individual's BMI over a decade. For males, the

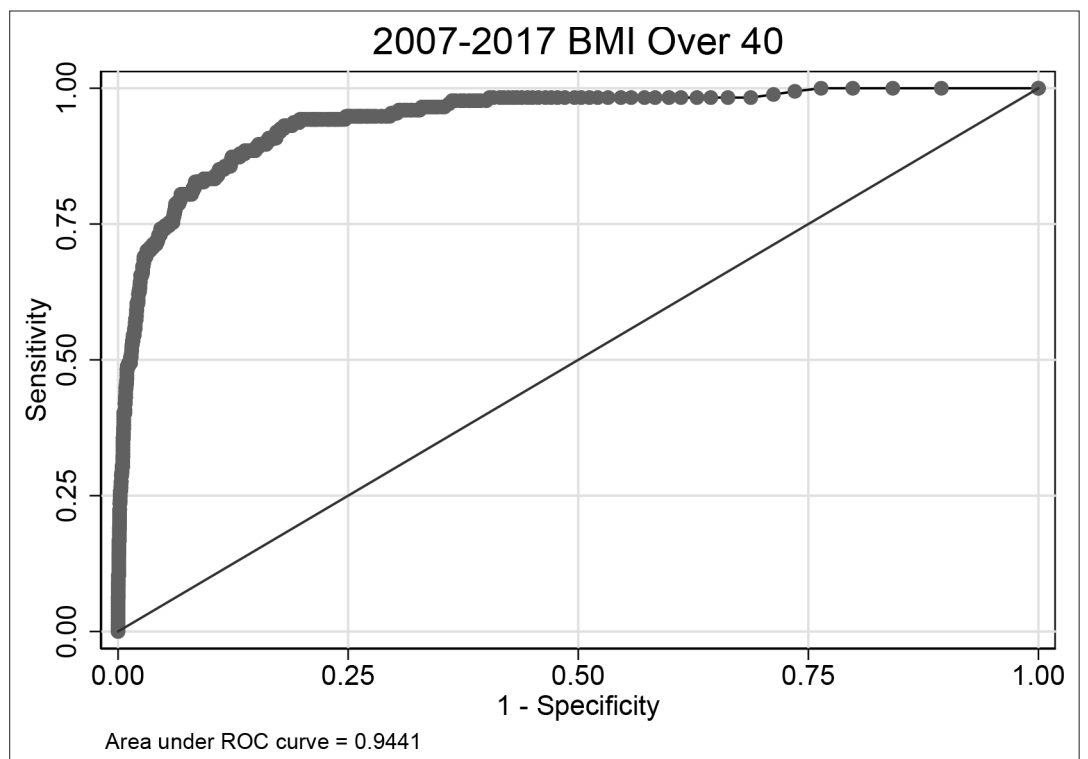


Figure 6. Receiver Operating Characteristic for BMI > 40 - Males

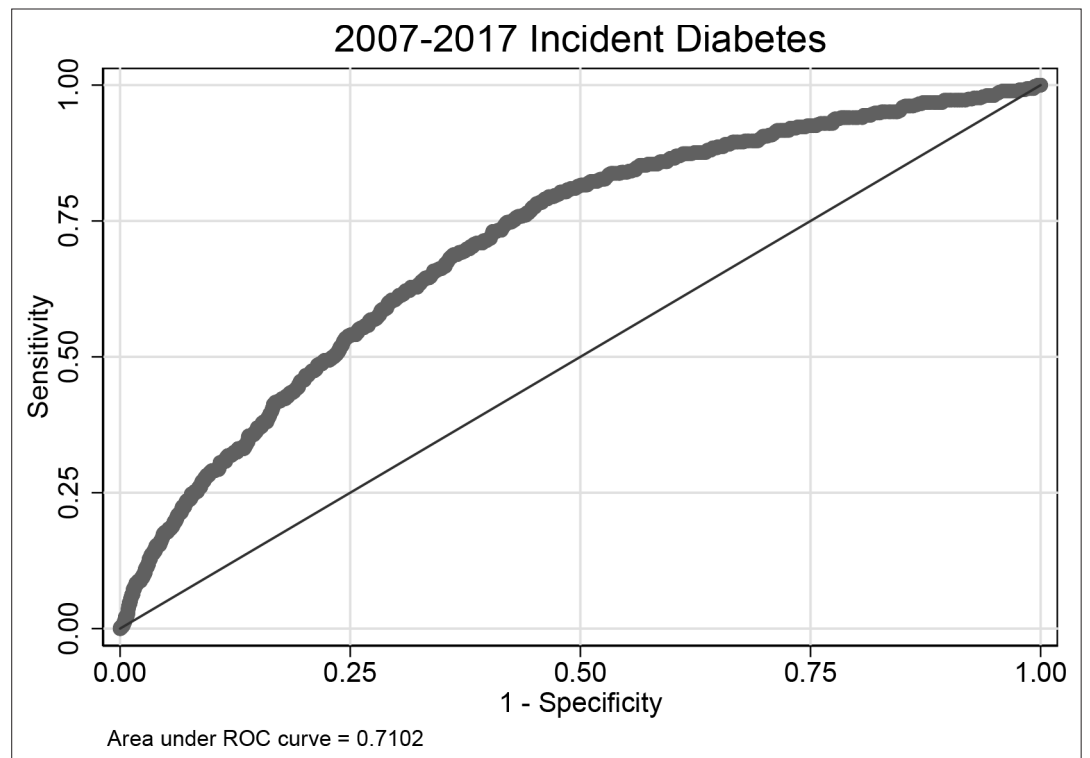


Figure 7. Receiver Operating Characteristic for Incident Diabetes - Females

AUC for the BMI under 25 classification is 0.90 (**Figure 4**), the BMI over 30 classification AUC is 0.92 (**Figure 5**), and the AUC is 0.94 for the over 40 classification test (**Figure 6**).

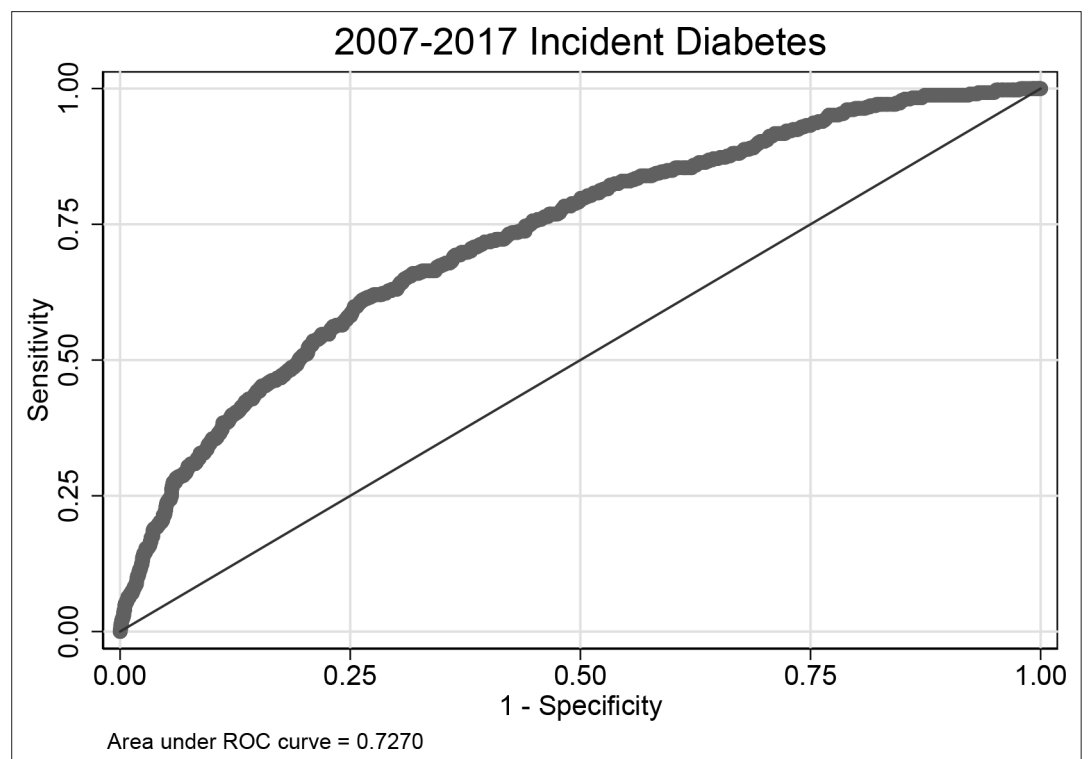


Figure 8. Receiver Operating Characteristic for Incident Diabetes - Males

Table 5. Host and external data comparison - 2007 PSID vs. 2007 BRFSS.

	Females		Males	
	PSID (2007)	BRFSS (2007)	PSID (2007)	BRFSS (2007)
BMI				
1st pctl	17.5 [17.2, 17.8]	17.6 [17.5, 17.7]	19.4 [18.8, 19.9]	19.2 [19.0, 19.3]
5th pctl	19.1 [19.0, 19.3]	19.5 [19.5, 19.6]	21.7 [21.5, 21.9]	21.5 [21.4, 21.6]
10th pctl	20.3 [20.1, 20.4]	20.5 [20.5, 20.6]	22.8 [22.6, 23.0]	22.6 [22.6, 22.7]
25th pctl	22.3 [22.2, 22.5]	22.7 [22.7, 22.7]	24.7 [24.5, 24.9]	24.6 [24.4, 24.7]
Mean	26.9 [26.7, 27.1]	27.2 [27.1, 27.2]	27.9 [27.8, 28.1]	28.1 [28.1, 28.2]
75th pctl	30.0 [29.7, 30.2]	30.2 [30.1, 30.2]	30.1 [29.9, 30.4]	30.6 [30.5, 30.7]
90th pctl	35.3 [34.7, 35.8]	35.5 [35.4, 35.6]	34.2 [33.8, 34.5]	34.5 [34.4, 34.6]
95th pctl	39.3 [38.6, 40.0]	39.2 [39.1, 39.4]	36.8 [36.3, 37.4]	37.4 [37.2, 37.6]
99th pctl	48.4 [46.2, 50.6]	47.9 [47.5, 48.4]	43.8 [42.2, 45.3]	45.0 [44.6, 45.5]
Diabetes prevalence	7.3 [6.5, 8.2]	9.2 [9.0, 9.4]	8.7 [7.8, 9.7]	9.9 [9.7, 10.2]

Notes: Confidence intervals for PSID and BRFSS reflect the complex survey design.

Predicting ten-year incidence of diabetes does not perform as well as predicting obesity or extreme obesity. The AUC for females is 0.71 (**Figure 7**) and the AUC for males is 0.73 (**Figure 8**). In addition to giving a sense of FAM’s classification ability, this validation method is also useful for assessing the predictive power in incorporating additional predictors.

4. External validity

The Behavioral Risk Factor Surveillance System is a survey designed to collect state-specific data on health risk behaviors, chronic diseases, and other health-related outcomes related to the leading causes of death and disability in the United States (**Centers for Disease Control and Prevention and others, 2018**). The sample size is large, with over 450,000 observations in 2017, and designed to be nationally representative. Like the PSID, BRFSS faces challenges with response rate. The self-reported measures in BRFSS are comparable to PSID. Diabetes status is asked in a similar manner (“Has a doctor, nurse, or other health professional EVER told you that you had diabetes?”). BMI is calculated from self-reported height and weight.

Before comparing FAM projections with BRFSS, it is valuable to show that PSID and BRFSS showed comparable BMI and diabetes summary statistics for those twenty-five and older in 2007. The twenty-five and older PSID respondents are the initial cohort in the validation exercise. Selected sample characteristics are presented in **Table 5** by gender. Female mean BMI is higher in BRFSS than PSID by 0.3 units and 0.2 units for males. At the 95th percentile, females in BRFSS are within 0.1 units of women in PSID, with men in BRFSS 0.6 units higher. At the 99th percentile the two samples are a bit further apart, with women in PSID 0.5 BMI units higher than BRFSS and men in PSID 1.2 units lower. Female PSID respondents report slightly lower BMI at the fifth through twenty-fifth percentiles. Diabetes rates also differ, with lower rates in PSID by 1.9 percentage points for women and 1.2 percentage points for men.

The BMI distributions for females 35 and older in FAM in 2007 and 2017 and BRFSS in 2007 and 2017 are shown in **Figure 9**. Examining the two BRFSS distributions, one sees the curve shifting to the right over the decade and the right tail pushing out. Qualitatively, FAM seems to capture this behavior. **Figure 10** illustrates the relative cumulative distribution between FAM and BRFSS in 2017 (left panel). The relative density functions are shown in the right panels. The Kullback-Leibler entropy is 0.018 [0.016, 0.020] and the median relative polarization index is 0.042 [0.035, 0.049]. **Table 6** compares these distributions in greater detail for 2017. Mean BMI for women is 0.5 BMI units lower in FAM projections than in 2017 BRFSS respondents. At the 95th percentile, FAM projections are 0.7 BMI units lower than BRFSS. At the 99th percentile, the distributions are further apart, with FAM 2.0 BMI units lower.

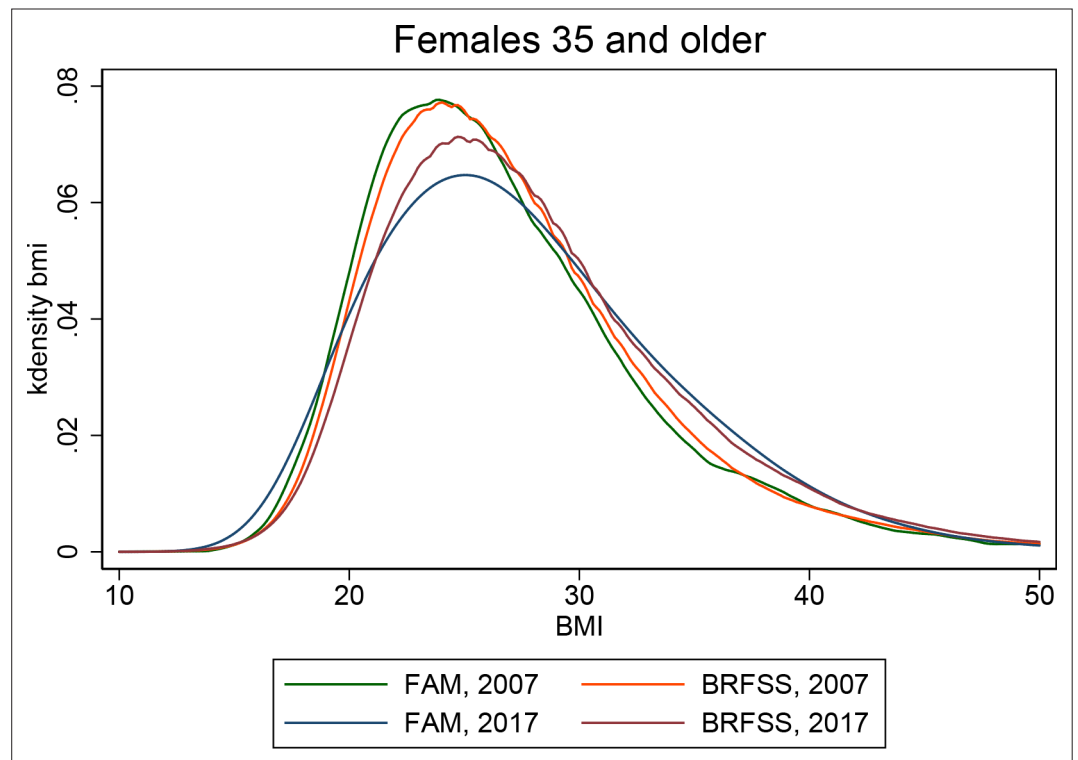


Figure 9. BMI distribution for those 35 and older - Females

The BMI distributions for males 35 and older are shown in **Figure 11**. After 10 years of FAM projection, the FAM distribution appears to have more dispersion than BRFSS. **Figure 12** shows the relative cumulative distribution and relative density functions. Entropy, as measured by the Kullback-Leibler

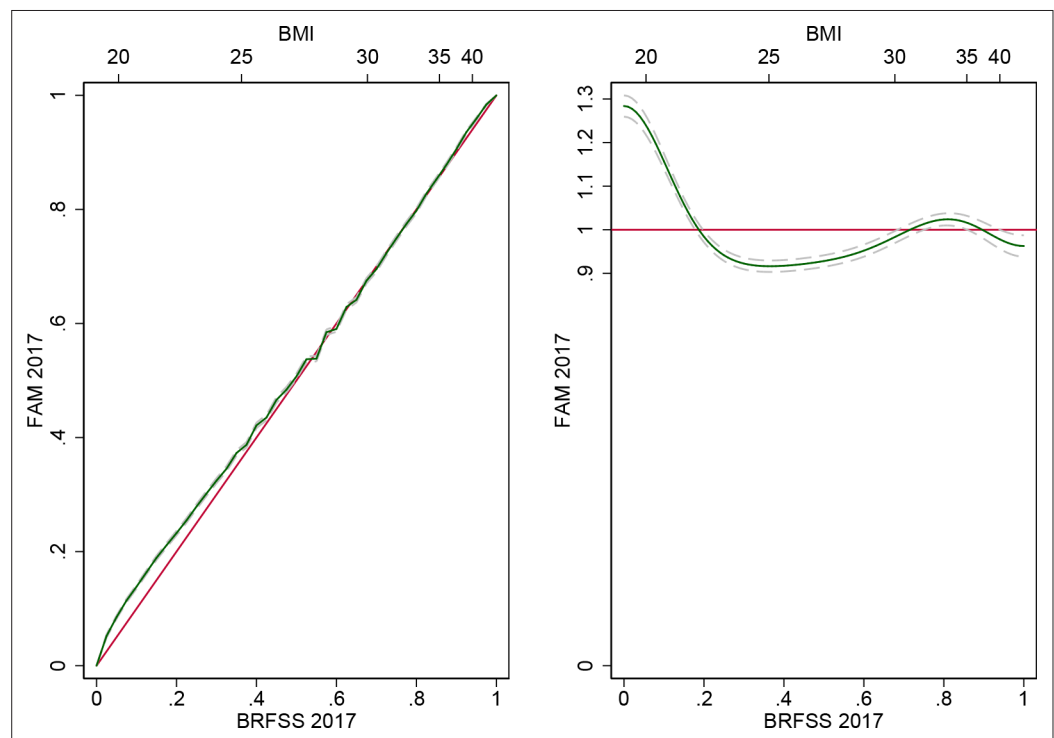


Figure 10. Relative cumulative distribution and density of BMI - Females

Table 6. External validation - 2017 FAM vs. 2017 BRFSS.

	Females		Males	
	FAM (2017)	BRFSS (2017)	FAM (2017)	BRFSS (2017)
BMI				
1st pctl	16.6 [16.2, 16.9]	17.7 [17.6, 17.8]	19.2 [19.0, 19.5]	18.8 [18.7, 19.0]
5th pctl	18.7 [18.5, 19.0]	19.8 [19.7, 19.9]	21.2 [21.0, 21.4]	21.6 [21.5, 21.7]
10th pctl	20.1 [20.0, 20.3]	21.0 [20.9, 21.0]	22.5 [22.2, 22.7]	23.0 [22.9, 23.0]
25th pctl	22.9 [22.8, 23.0]	23.4 [23.4, 23.5]	24.9 [24.7, 25.0]	25.1 [25.1, 25.1]
Mean	27.7 [27.6, 27.9]	28.2 [28.1, 28.3]	28.5 [28.3, 28.7]	28.8 [28.7, 28.9]
75th pctl	31.5 [31.3, 31.8]	31.6 [31.5, 31.8]	31.5 [31.2, 31.8]	31.6 [31.5, 31.7]
90th pctl	36.7 [36.3, 37.0]	36.9 [36.7, 37.1]	35.3 [34.8, 35.7]	35.9 [35.7, 36.0]
95th pctl	40.0 [39.4, 40.5]	40.7 [40.5, 41.0]	37.7 [37.1, 38.2]	39.0 [38.8, 39.3]
99th pctl	46.9 [45.5, 48.2]	49.9 [49.3, 50.5]	42.4 [41.3, 43.6]	46.8 [46.2, 47.4]
Diabetes prevalence	12.4 [11.5, 13.3]	14.0 [13.7, 14.4]	14.9 [13.6, 16.3]	15.7 [15.3, 16.1]

Notes: Confidence intervals for FAM reflect 50 sets of bootstrapped transition models, each simulated 100 times. Confidence intervals for BRFSS reflect the complex survey design.

divergence is 0.015 [0.013, 0.017] and median relative polarization is .018 [0.011, 0.025]. The 2017 distributions are compared in **Table 6**. Here, we see that mean BMI in FAM is 0.3 BMI units below BRFSS. At the 75th percentile FAM is 0.1 BMI units below BRFSS, FAM is 1.3 BMI units lower at the 95th percentile, and the distributions are 4.2 BMI units apart at the 99th percentile.

Diabetes prevalence for the 35 and older population is also shown in **Table 6**. FAM is 1.6 percentage points lower than BRFSS for women and 0.8 percentage points lower for men. This is consistent with the initial differences observed in 2007 for the twenty-five and older PSID compared to BRFSS.

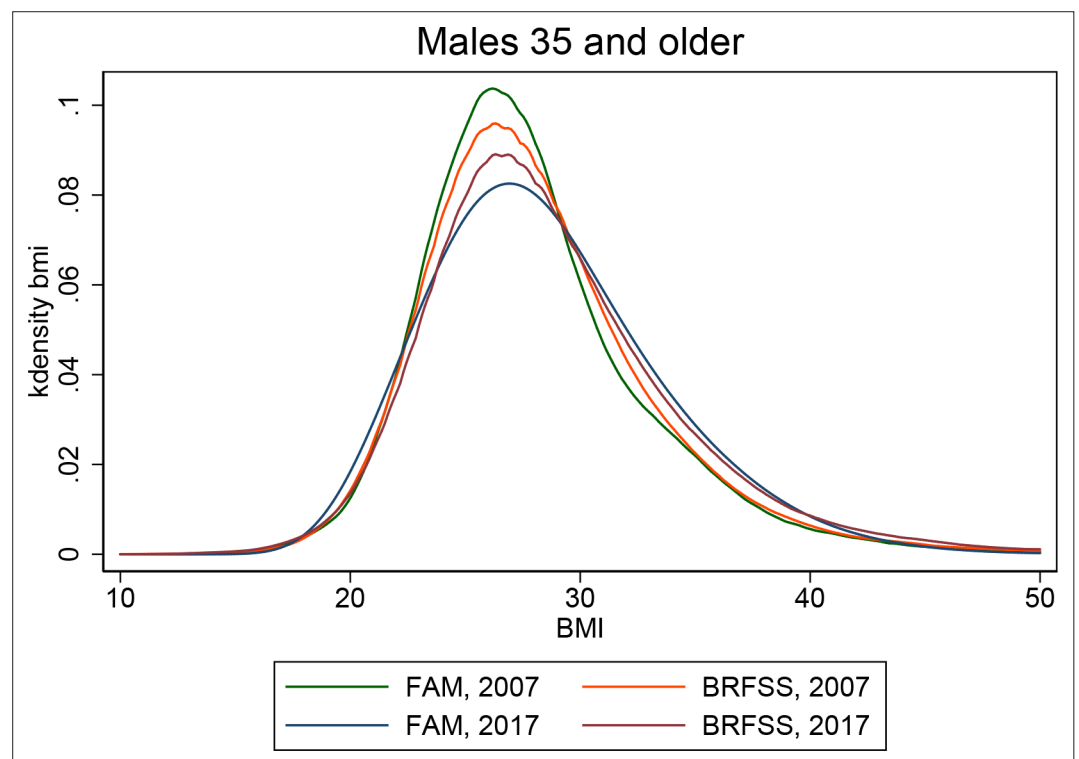


Figure 11. BMI distribution for those 35 and older - Males

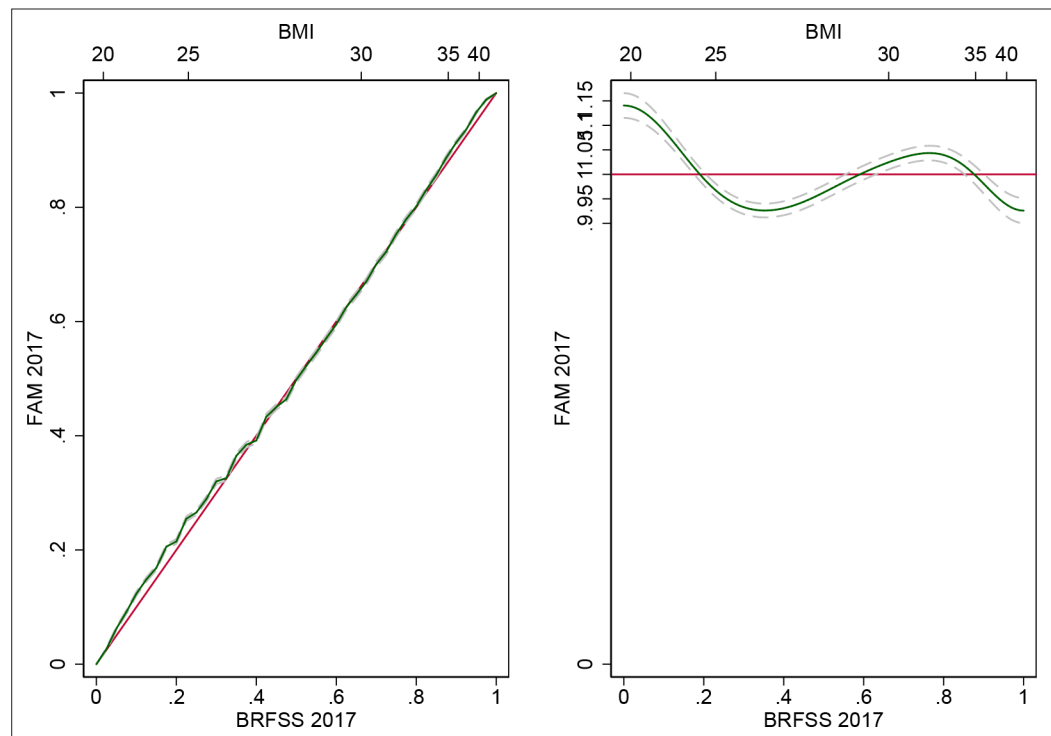


Figure 12. Relative cumulative distribution and density of BMI - Males

5. Discussion

In estimating transition models from 1999-2007, then applying them to simulate 2007-2017, the analysis shown here is analogous to a period life table approach. The key assumption is that the 1999-2007 observed transitions will continue to hold through the forecasting period. For the outcomes of interest, one can imagine situations in which this assumption is violated, such as changes in diagnostic procedures for diabetes or the adoption of public health campaigns targeting obesity. Consequently, deviations from what was observed can be driven by both simulation limitations and real world changes.

In projecting BMI, the 2017 mean BMI in FAM is slightly higher than 2017 PSID (0.1 BMI units women, 0.2 for men), while slightly lower than 2017 BRFSS in means (0.5 BMI units for women, 0.3 for men). For the right tail of the distribution, 2017 FAM BMI distributions are lower than both the PSID and BRFSS at the 95th percentile, much lower at the 99th. This suggests that FAM performed well for much of the distribution, but not for extreme cases. The ROC analyses for BMI are likely driven by the persistence of BMI. For example, those with a BMI over 40 are likely to have a BMI over 40 ten years later.

For consumers of the model interested in BMI, this validation exercise suggests that projections of many groups of interest are plausible, such as the fraction overweight, obese, or severely obese. However, the distribution of BMI within the severely obese category in FAM under-predicted the BMI values observed in both PSID and BRFSS by two to four BMI units. If the actual BMI values are of interest for those cases, then the BMI modeling approach would merit attention, perhaps including calibrating to an external source.

Diabetes rates projected with FAM are lower than both 2017 PSID and 2017 BRFSS. The lower rates in 2007 PSID compared to 2007 BRFSS are preserved in FAM. Individual assignment of incident diabetes does not perform as well as models with more information, as shown below. The FAM diabetes model controls for age, race, education, childhood characteristics, smoking, exercise, and BMI. This yields AUC statistics of 0.71 for twenty-five and older women, 0.72 for twenty-five and older men. Notably, the transition model does not currently incorporate family history, pharmaceutical use, biometrics (blood pressure, waist measurements), biomarkers (fasting glucose, cholesterol), or diet. In a UK sample of 40 to 79 year olds, a risk model based on self-reported data on medication, BMI,

family history, age, nutrition, and exercise yielded an AUC of 0.763 for predicting incident diabetes in 4.6 years (*Simmons et al., 2007*). In a US-based study of 25 to 64 year olds, the San Antonio Heart Study, incidence of diabetes was assessed in several different models incorporating a variety of clinical tests. These models had AUC statistics ranging from 0.775 to 0.859. The clinical multivariate models incorporated age, sex, ethnicity, fasting glucose level, systolic blood pressure, HDL cholesterol, BMI, and family history of diabetes (*Stern et al., 2002*). In a Finnish study of 10-year incidence of drug-treated diabetes amongst 35 to 64 year old individuals, a model adjusting for age, BMI, waist circumference, use of blood pressure medication, a history of high blood glucose (including those with diabetes, but not taking medication), physical activity, and diet yielded an AUC of 0.860 (*Lindström and Tuomilehto, 2003*).

For consumers of the model interested in diabetes, this validation exercise suggests that projections of diabetes follow the same trend as in BRFSS. The initially lower rates in PSID are maintained over the decade of projection. If possible, additional information on family history, medications, diet, or other clinical measures would likely improve the specificity and sensitivity of FAM. The AUC statistics for models that include additional information give a sense of how well FAM might perform with more information.

Overall, these validation exercises are reassuring. The PSID host data compare well with BRFSS. Trends, such as the shift right in BMI and the expansion of the right tail of the BMI distribution, are captured with fairly simple transition models that do not assume a temporal trend. Individual-level predictions are concordant for various classifications of BMI, likely due to the persistence of BMI for individuals, and diabetes incidence classification is reasonable given the lack of clinical information.

ORCID iD

Bryan Tysinger  <https://orcid.org/0000-0002-8497-3664>

Acknowledgements

Not applicable.

Funding

This work was supported by the National Institute on Aging of the National Institutes of Health under award numbers P30AG024968 and R01HD087257.

Conflict of Interest

No competing interests reported.

Data and Code Availability

The data used in FAM are available for scientific research upon registration.

1. Panel Study of Income Dynamics is available at <https://psidonline.isr.umich.edu/>
2. Behavioral Risk Factor Surveillance System is available at <https://www.cdc.gov/brfss/index.html>
3. Health and Retirement Study is available at <http://hrsonline.isr.umich.edu/>
4. Medical Expenditure Panel Survey is available at <https://meps.ahrq.gov/mepsweb/>
5. Medicare Current Beneficiary Survey is available at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS>
6. National Health Interview Survey is available at <https://www.cdc.gov/nchs/nhis/index.htm>
7. National Health and Nutrition Examination Survey is available at <https://www.cdc.gov/nchs/nhanes/index.htm>
8. AmericanCommunitySurvey is available at <https://www.census.gov/programs-surveys/acs>
9. CensusPopulation Projections is available at <https://www.census.gov/programs-surveys/popproj.html>

The source code for FAM is available via a publicly accessible Subversion code repository at <https://schweb.lahrc.lahealthresearchcloud.org/svn/PublicFutureElderlyModel/trunk>. It is suggested that users contact Bryan Tysinger for guidance in getting started.

References

- Centers for Disease Control and Prevention and others.** 2018. *Behavioral risk factor surveillance system overview: Bfss 2017*.
- Clair PS**, Gaudette É, Zhao H, Tysinger B, Seyedin R, Goldman DP. 2017. Using self-reports or claims to assess disease prevalence: its complicated. *Medical care* **55**:782.
- Fryar C**, Carroll M, Ogden C. 2019. *Prevalence of overweight, obesity, and severe obesity among adults age 20 and over: United States, 1960-1962 through 2015-2016*. Bethesda, MD: National center for health statistics 2018.
- Lindström J**, Tuomilehto J. 2003. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* **26**:725–731. DOI: <https://doi.org/10.2337/diacare.26.3.725>
- McGonagle KA**, Schoeni RF, Sastry N, Freedman VA. 2012. The panel study of income dynamics: overview, recent innovations, and potential for life course research. *Longitudinal and life course studies* **3**. DOI: <https://doi.org/10.14301/lcs.v3i2.188>
- Menke A**, Casagrande S, Geiss L, Cowie CC. 2015. Prevalence of and trends in diabetes among adults in the United States, 1988-2012. *JAMA* **314**:1021–1029. DOI: <https://doi.org/10.1001/jama.2015.10029>
- Pandya A**, Sy S, Cho S, Alam S, Weinstein, Gaziano TA. 2017. Validation of a cardiovascular disease policy micro-simulation model using both survival and receiver operating characteristic curves. *Medical Decision Making*:0272989X17706081.
- Simmons RK**, Harding A-H, Wareham NJ, Griffin SJ, Team E, EPIC-Norfolk Project Team. 2007. Do simple questions about diet and physical activity help to identify those at risk of type 2 diabetes? *Diabetic Medicine* **24**:830–835. DOI: <https://doi.org/10.1111/j.1464-5491.2007.02173.x>
- Stern MP**, Williams K, Haffner SM. 2002. Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? *Annals of Internal Medicine* **136**:575–581. DOI: <https://doi.org/10.7326/0003-4819-136-8-200204160-00006>