

**Estimating the discrete probability distribution
of the age characteristic of Veteran populations using
SAS[®], SAS/OR[®] and SAS Simulation Studio[®] for use
in population projection models**

Michael C. Grierson,
The Department of Veterans Affairs,
Washington, DC

The views expressed in this paper are those of the author and do not reflect the official policy or position of the Department of the Veterans Affairs, or the U.S. Government.

ABSTRACT

This paper examines the discrete probability distribution of the age characteristic of the Veteran population in the US. The basis of the examination is the American Community Survey¹ (ACS) and the Social Security Administrations (SSA) period life table for males in 2007.² The ACS is stratified to efficiently represent population elements including the population of Veterans. The current Veteran population discrete probability distribution for age is examined as a starting point. The survey also supports examining the expected future age of Veteran discrete probability distributions. The paper will show how to use an efficiently designed stratified survey (the 1 year ACS) to support a simulation that moves that population into the future and provides a basis for estimating future population age distributions and their consequent impact on organization cost and budget planning. Combining the processing of stratified survey data with Operations Research tools (like the discrete event simulator) and supporting subsequent analysis of results is a combined capability that is not matched elsewhere.

INTRODUCTION

The broad objectives of the American Community Survey 2010 are to provide the latest statistics on American Communities³. The Census Bureau provides ACS use examples and documents insights about the surveys design⁴. The ACS is managed by the American Community Survey office of the Census Bureau. The Census design was to use a stratified survey to capture data about American Communities that include Veterans, the Veteran utilization of Veterans Health Administration (VHA) services, the Veterans latest period of service and a few other Veteran attributes (like year of birth). The ACS supports comparisons between Veteran populations and the greater Communities where they reside. The stratified survey approach is also a cost efficient way to minimally burdened Communities while capturing many of those population characteristics of interest for Federal Government planning. The ACS stratified survey and the survey procedures of SAS[®] will be used to determine estimates of Veteran population characteristics that will support building a population simulation using SAS Simulation Studio[®]. The simulation will be used to project the population age distribution forward in time.

SURVEY DATA

The ACS is stratified to efficiently and effectively represent the particulars of the sampled US population. Stratified surveys randomly sample homogenous sub-populations (rural vs. urban for example) and combines the sub-populations using weights to correct the stratifications (randomly sampled sub-populations) back to a representative sample of the combined population (here the US population). To begin working the ACS data, the 'survey' procedure, SURVEYFREQ will be used to estimate the population distribution of the age characteristic and to assess the impact of age on VA service profiles. Veterans have a differing VA services profile as they age. For example, they are more likely to use GI Bill benefits right when they leave military service then at later periods in their lives, largely because those benefits are no longer available beyond 10 or 15 years after their last active duty date. Some Veterans are more likely to prefer to include VA hospitals in their health care choices as they age. The SAS survey procedures also support regression analysis via the SURVEYLOGISTIC procedure. The SURVEYLOGISTIC procedure can also be used to assess the services decision dependent variables (like the use of GI Bill benefits) relative to the independent variables of age or age category.

¹ The 2010 1 year American Community Survey results at http://www2.census.gov/acs2010_1yr/pums/unix_pus.zip

² <http://www.ssa.gov/oact/STATS/table4c6.html>

³ http://www.census.gov/acs/www/guidance_for_data_users/comparing_2010/ and http://www.census.gov/acs/www/guidance_for_data_users/estimates/

⁴ http://www.census.gov/acs/www/guidance_for_data_users/handbooks/

The SURVEYFREQ procedure shown in Figure 1 below is used to estimate the age distribution of the Current Veteran population.

```

libname pus "C:\somedir";
libname frmlib "C:\somedir";
%let graphpath "C:\somedir";
option fmtsearch=(frmlib work);
data pus.veterans;
  set pus.pus2010;
  if mil < 2 or mil > 3 then delete;
run;
proc surveyfreq data=pus.veterans;
  table agep;
  ods output OneWay=pus.freqout;
  weight pwgtp;
run;
data pus.vetagepdf;
  set pus.freqout;
  vageprob=round(Percent/100,0.000001);
  if vageprob > 0.999 then delete;
run;
goptions reset=all;
ods html gpath=&graphpath;
proc sgplot data=pus.vetagepdf;
  title "Current Veterans";
  xaxis min=0 max=105;
  series x=Agep y=vageprob ;
run;

```

Figure 1. SURVEYFREQ procedure

Note that the stratified survey parameter of weight (pwgtp) is critical to handling the stratifications of the survey and the model is outputting data to a dataset using the SAS/ODS® output facility. The age is calculated from the agep (labeled PUMS age in the dataset) survey variable. The vageprob variable uses the Percent column (which is the percent of weighted counts) from the OneWay output to show the probabilities (percents are 100 times probabilities) of Veterans for each age. The output of this procedure looks something like as shown in figure 2 below.

Probability of Vet Age, for 2010

Current Veterans 12:23 Saturday, March 24, 2012 182

The SURVEYFREQ Procedure

Data Summary

Number of Observations	244785
Sum of Weights	21775185

Current Veterans 12:23 Saturday, March 24, 2012 183

The SURVEYFREQ Procedure

PUMS Age

AGEP	Weighted Frequency	Std Dev of Frequency	Wgt Freq	Std Err of Percent	Percent
17	20	1514	368.69390	0.0070	0.0017
18	65	7738	1211	0.0355	0.0056
19	162	18569	1781	0.0853	0.0082
20	274	37063	2735	0.1702	0.0126

21	379	49785	3239	0.2286	0.0149
22	464	57176	3233	0.2626	0.0148
23	592	76410	3943	0.3509	0.0181
24	788	98466	4368	0.4522	0.0200
25	870	108777	4541	0.4995	0.0208
26	992	125594	4908	0.5768	0.0225
27	1142	138576	5056	0.6364	0.0232
28	1208	143138	5079	0.6573	0.0233

*
*
*

Figure 2. SURVEYFREQ output sample

Note that the output shows the sampling stratifications and the resulting weighted frequencies. The column of interest for the distribution is Percent. This column and the SGPLOT procedure can be used to produce the probability distribution shown in the graph in Figure 3 below.

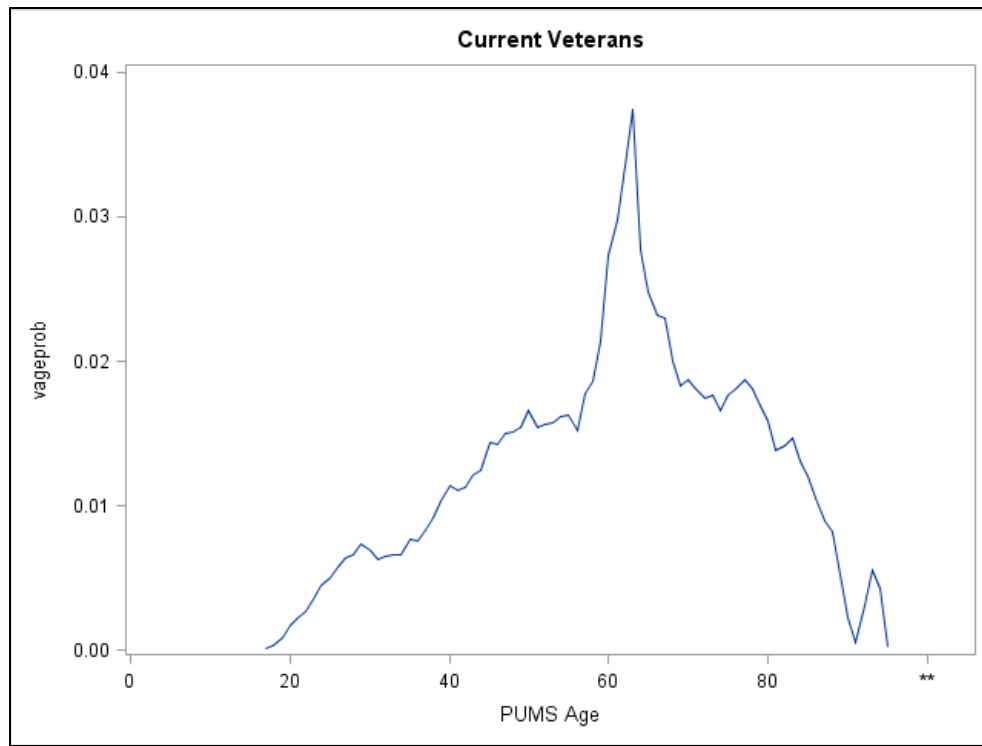


Figure 3. Probability distribution of Current Veteran age

A similar strategy can be used to calculate the age when the military service member became a Veteran (*vetage* in Figure 1). The survey asked the question, "When did you serve on active duty in the U.S. Armed Forces"⁵? The PUMS Age and the last year of possible active duty (estimated from the categorized answers) can be used to estimate the age of becoming a Veteran (called *vetage* in the program). The SGPLOT procedure can be applied as before to produce an estimate of the age probability distribution for Future Veterans which is shown below in Figure 4. Additional investigation reveals that the pattern shown in Figure 4 is consistent for most age generations. The conclusion is that when the Department of Defense (DoD) creates a Veteran, they enter the Veteran population while at the ages represented in Figure 4. This will be assumed as generally valid for future Veterans entering the population (it is valid for past populations for example the decennial census 1990 and 2000).

⁵ Question 27 of the 2010 ACS questionnaire can be found here, <http://www.census.gov/acs/www/Downloads/questionnaires/2010/Quest10.pdf>

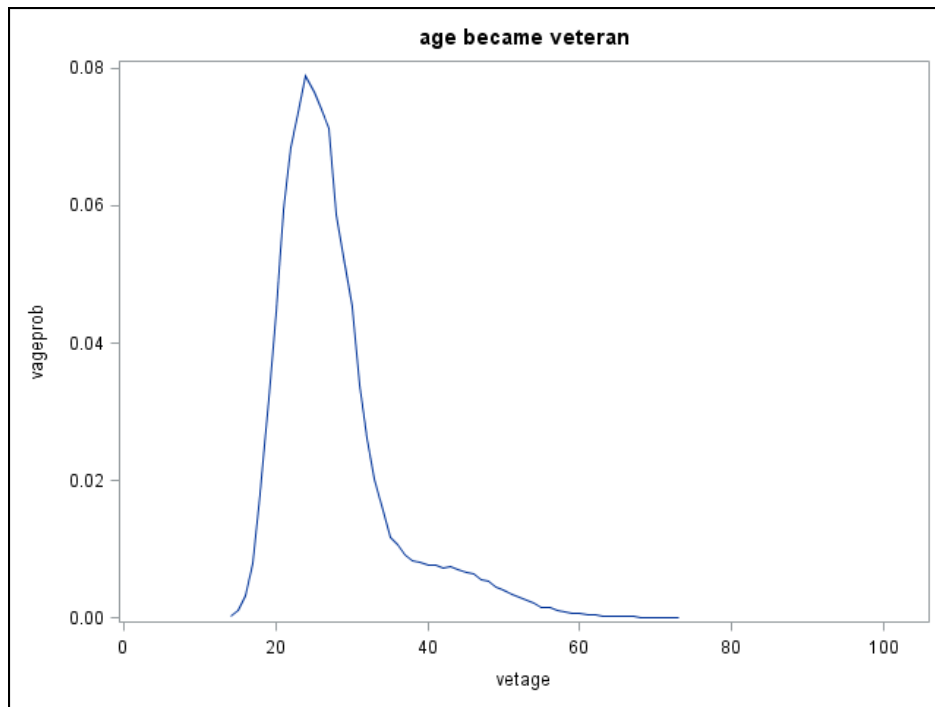


Figure 4. Probability distribution of age when became a Veteran

The age distributions will be converted to years of expected remaining life distributions, primarily to project when the VA obligation to a Veteran ends but also to support projecting future service delivery obligations for services. The current population is as of 2010 and future Veterans will enter the population periodically each subsequent year. Knowing when the VA obligation ends can be determined with the current Age and the expected life tables from the SSA. The probability distributions of the Expected years to death for Current and Future populations are shown below in Figure 5.

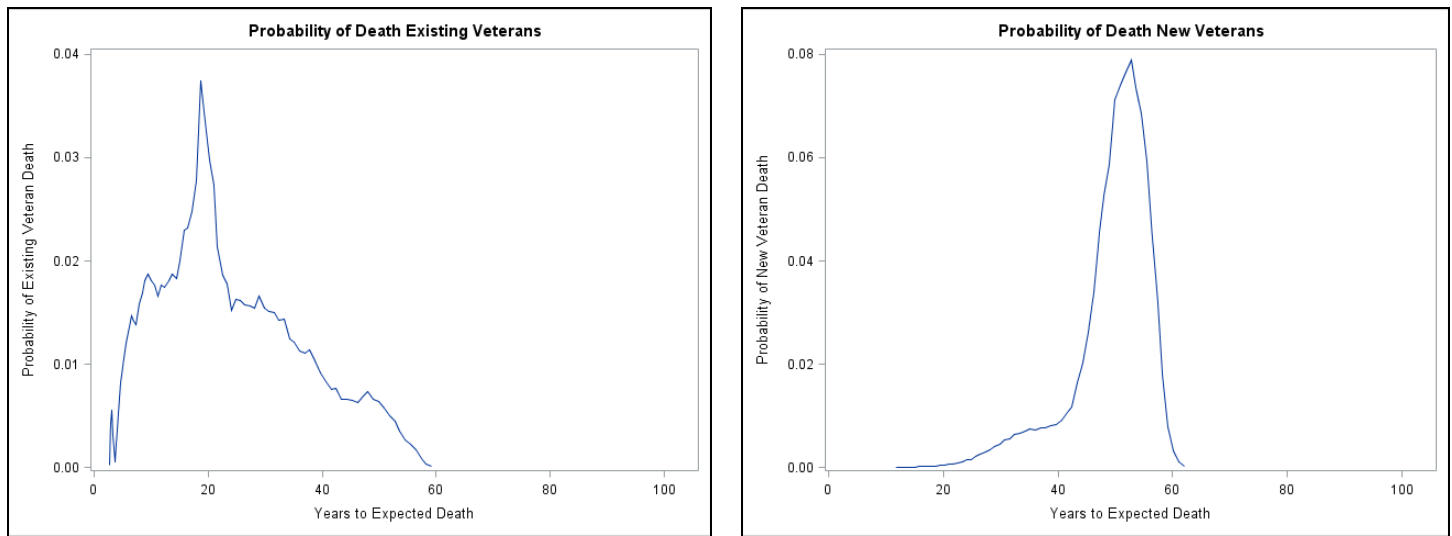


Figure 5. Distribution of Expected years to death for Current and Future Veterans

A figurative model of the process flow is shown below in Figure 6, where the graphic shows the current Veteran population distribution is put into the model when it initiates, and the new Veterans being added to the model every year. The Social Security Administrations Life Tables are used to estimate the event of the Veterans aging out of the population. By using Discrete Probability Distributions, we can vary the amount of new Veterans entering the population. Annual attrition from the military is an often debated and closely watched number, but a good model will support a variety of choices to examine the effect on the VA of differing attrition from DoD scenarios.

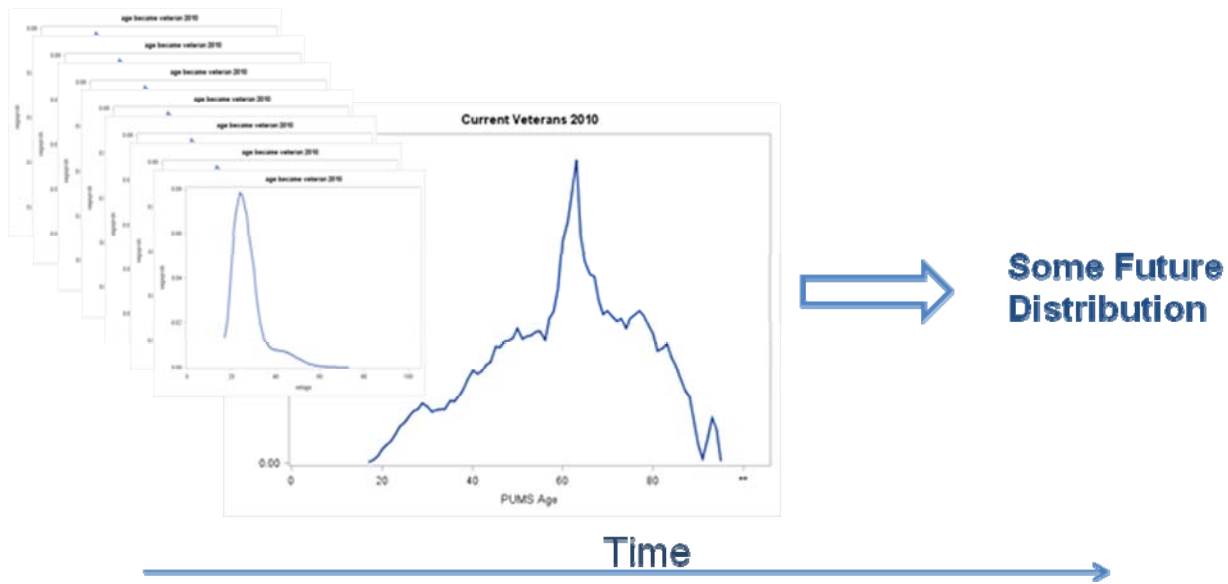


Figure 6. A process flow model for simulating population changes

POPULATION SIMULATION

The probability distributions shown previously, will be used to simulate the Veteran population as it moves forward in time. The 2010 ACS (1 year) provides the an estimate of the current Veteran population at 22.3 million. The DoD reports attrition to vary between 180,000 and 250,000 per year (see Table 1 below) and these numbers will be used for new Veteran creation rates. The simulation uses data from the current population age distribution and current population size, the new Veteran distribution and DoD attrition rates (averaged) and the expected life table data. The simulation will be used to capture living Veteran attributes for future years as a relatively smaller new Veteran population arrives yearly and as the larger Current Veterans age out of the population. An initial Model is shown below in Figure 7.

Some key features of this model are the use of the probability distributions shown in Figure 5 to drive the aging out process. The initial batch size is 22000 to represent one 'entity' per 1000 Veterans. 'Person' entities are created and given the attribute of expected life remaining from the dataset used in the Figure 5 graph (left hand side). In the model above, the FutureBatchSize is set to 0. The DeathAgeCV (in this model, FV is short for Future Veteran and CV is short for Current Veteran) number generator block actually has the data used to make the SGPLOT (as stored in a dataset) as a source. Note the similarity between the Histogram in Figure 6 and the left graphic in Figure 5.

To check the distribution of the Future Veterans, the batch size from the current Veteran distribution is set to zero and one batch of 22000 entities can be run through the DeathAgeFV number generator. The results are shown in Figure 8. The future population in Figure 7 is artificially generated to demonstrate the future distribution. These entities won't all come in one large batch, but will 'drip' into the population in a small batch per each future year. Table 1 below Figure 8 shows that the separation rates for the services is decreasing over time.

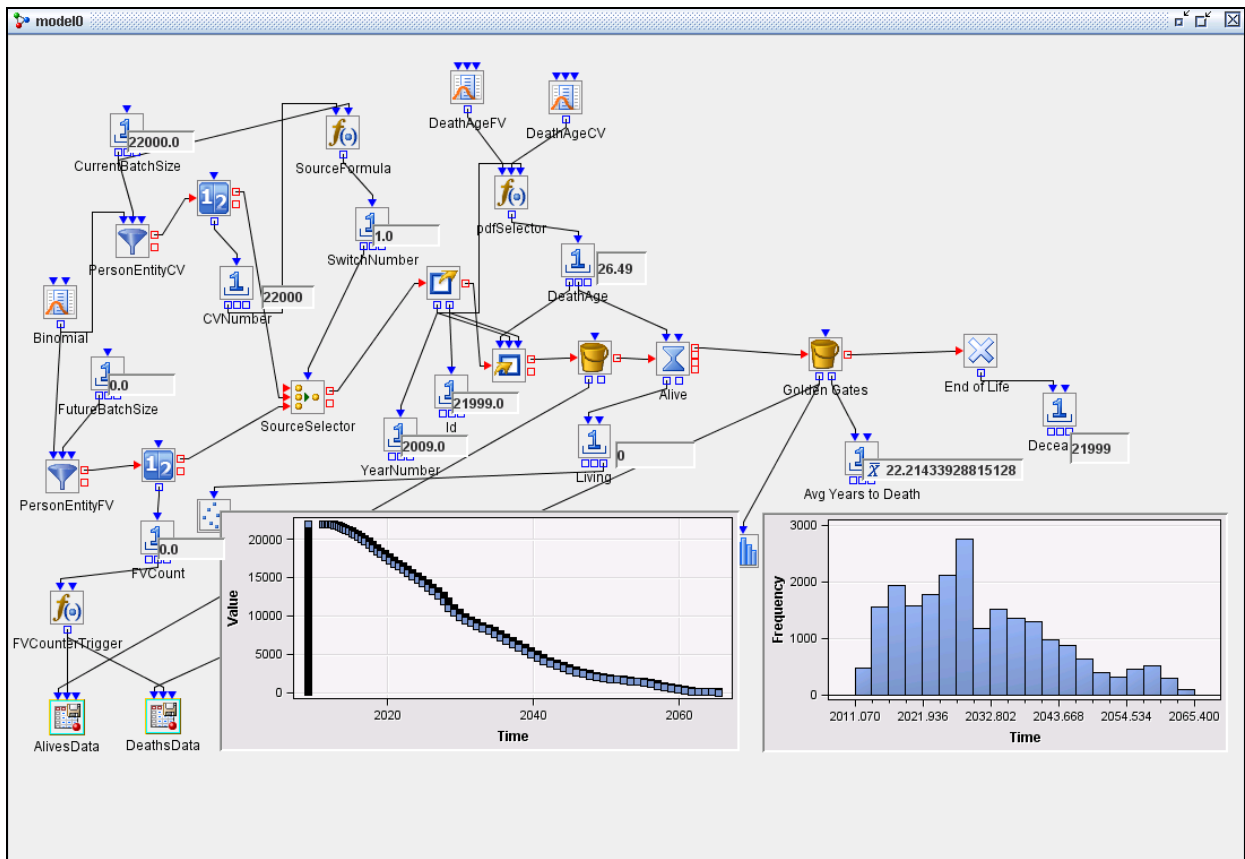


Figure 7. Initial SAS Simulation Studio® model showing current population aging out.

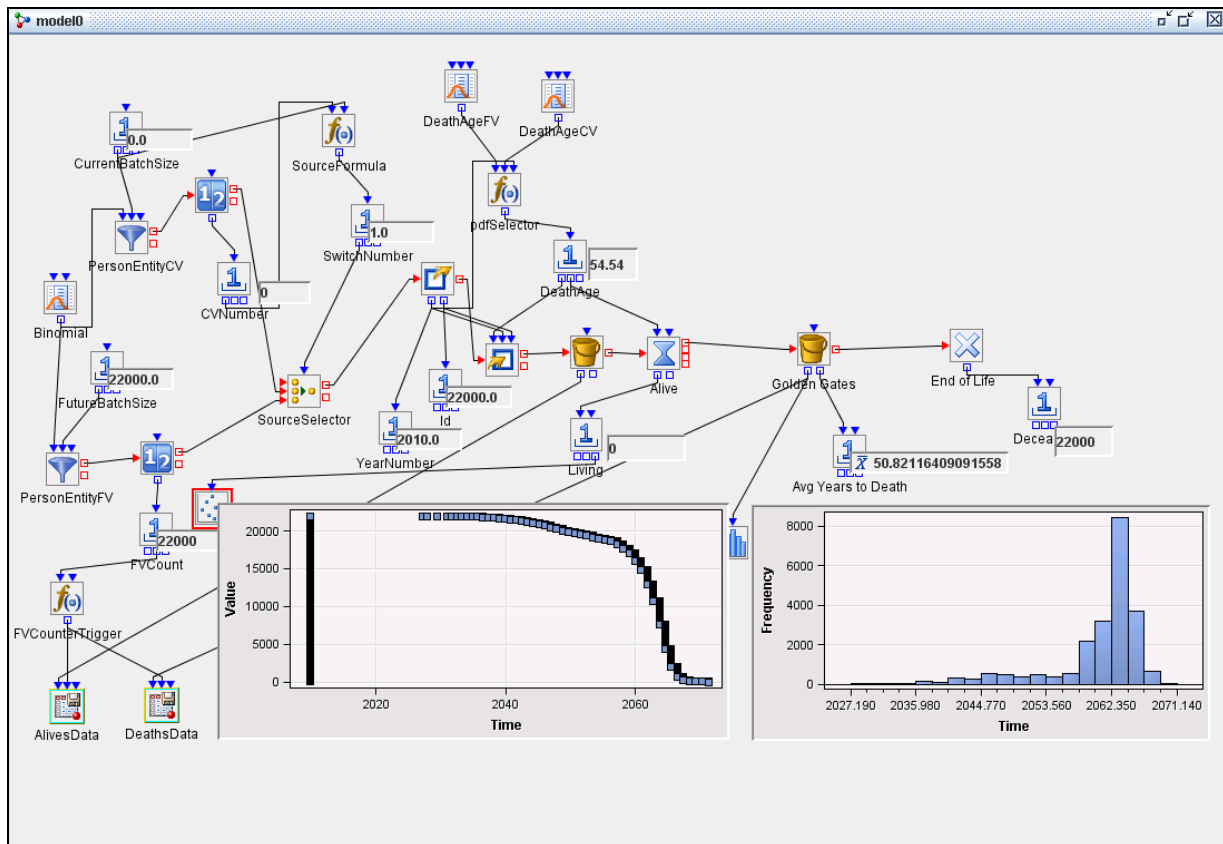


Figure 8. Initial SAS Simulation Studio® model showing ONE future population life aging out.

Table 1. Attrition from Active Duty Status⁶

FY 2005	Retired	Separated	Total
Army	10,989	74,296	85,285
Navy	10,248	71,572	81,820
Marine Corps	2,799	29,464	32,263
Air Force	12,845	34,551	47,396
Coast Guard	908	3,347	4,255
Total	37,789	213,230	251,019

FY 2006	Retired	Separated	Total
Army	10,676	60,357	71,033
Navy	10,106	43,531	53,637
Marine Corps	2,797	31,611	34,408
Air Force	11,304	29,498	40,802
Coast Guard	991	3,497	4,488
Total	35,874	168,494	204,368

FY 2007	Retired	Separated	Total
Army	9,940	62,874	72,814
Navy	9,338	44,743	54,081
Marine Corps	2,280	31,503	33,783
Air Force	12,911	35,444	48,355
Coast Guard	1,035	3,023	4,058
Total	35,504	177,587	213,091

FY 2008	Retired	Separated	Total
Army	9,662	61,921	71,583
Navy	8,985	39,248	48,233
Marine Corps	2,271	27,292	29,563
Air Force	9,257	26,942	36,199
Coast Guard	957	2,581	3,538
Total	31,132	157,984	189,116

FY 2009	Retired	Separated	Total
Army	10,697	61,131	71,828
Navy	8,471	34,908	43,379
Marine Corps	2,223	28,407	30,630
Air Force	8,924	22,040	30,964
Coast Guard	967	2,619	3,586
Total	31,282	149,105	180,387

Figure 9 below shows the results of periodically adding 200 entities per year into the simulation. Notice that a steady state condition shows a Veteran population of about 10.2 million Veterans. By periodically adding only 180 entities per year, the Veteran population is down to 9.2 million at steady state. Assumptions about how many Veterans are created each year will change with time and with the size of the DoD active duty population. As these changes occur, the model can be adjusted to support more accurate forecasting.

⁶ <http://www.stripes.com/news/as-separation-rates-plummet-military-getting-more-selective-1.96679>

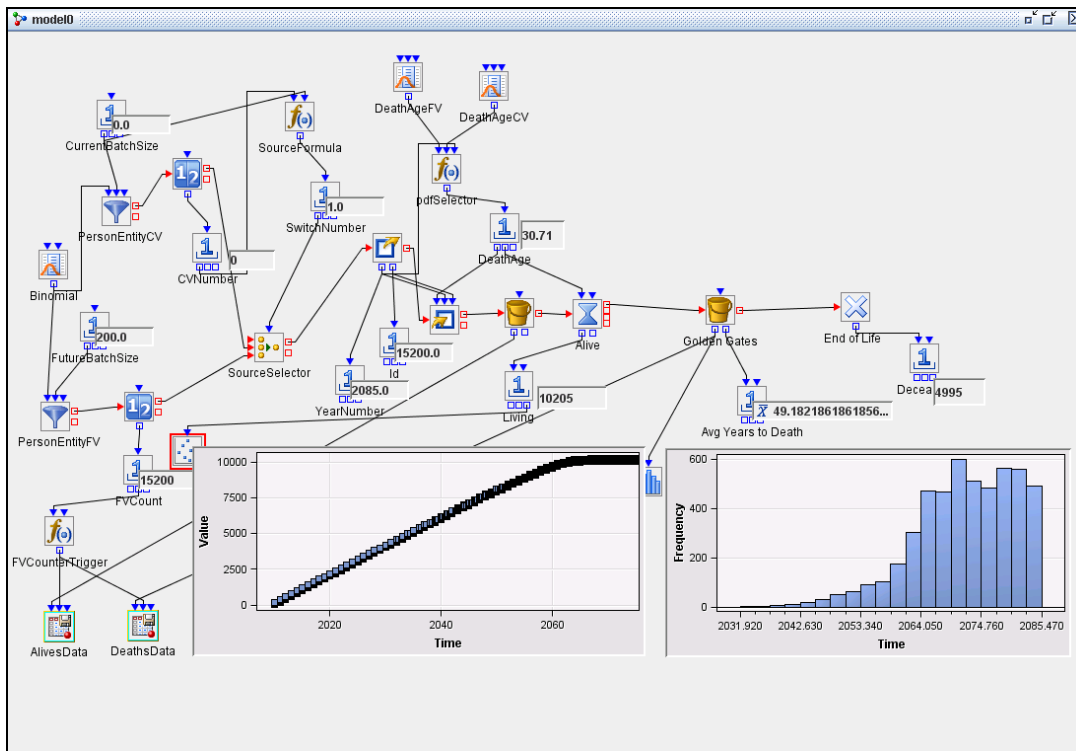


Figure 9. Initial SAS Simulation Studio® model showing many future populations aging to a steady population state.

By setting both batch sizes (current and future) to be representative of actual, we can show an estimate of the future Veteran population. The current and future contributions to a population results are shown in Figure 10 below. This model run gives us a general feel for short term Veteran population sizes, but the more practical interest is the living Veteran population at a future point in time and its age characteristics.

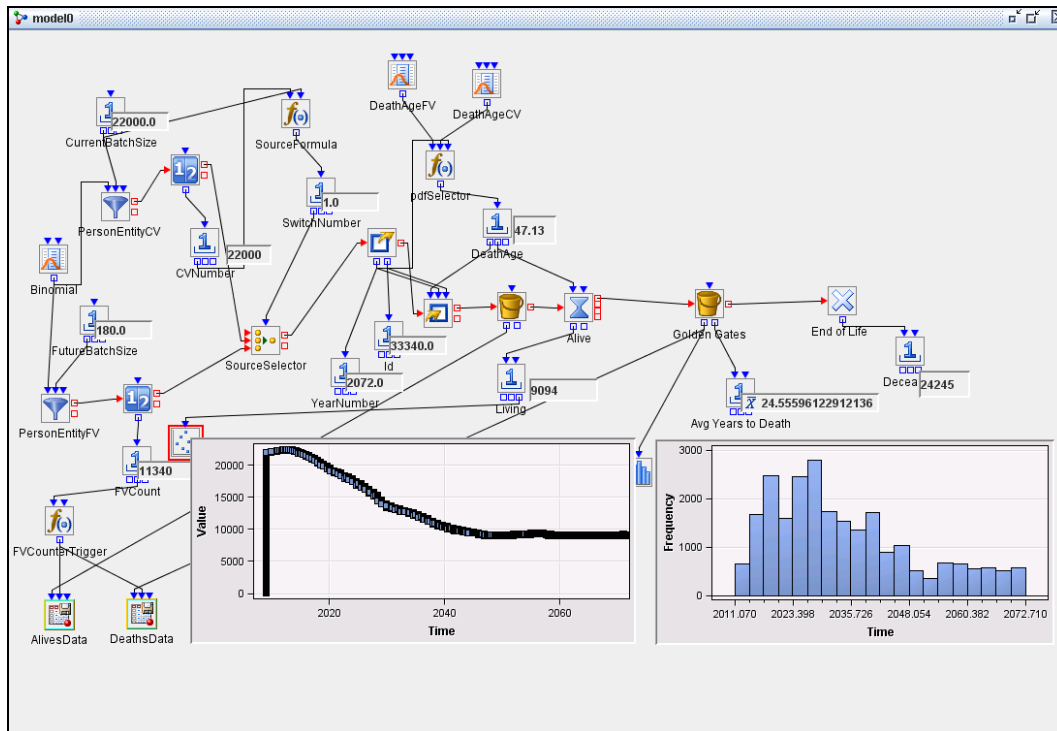


Figure 10. SAS Simulation Studio® model showing Veteran populations aging to a steady population state.

THE LIVING

To capture the age of the Veterans who are alive in any given future year, a trigger is created to write out to datasets all the entity attributes that have not reached their end of life time, and all the entity attributes that have been created up to a point in time. Two datasets are created, one of entity attributes that enter the 'Alive' block, and a dataset of entity attributes after they 'leave the Alive block' or expire. By subtracting the expired records from the non-expired records, a snapshot of the 'Living' vets at some point in time (usually triggered by an 'FV' entity count) can be estimated.

For example, if the FVCounterTrigger is set for a value of 3780 Future entities, the snapshot of data will be captured 21 simulation years after the start of the simulation. This should provide a snapshot of 'living' entities for the year 2030 based on Life Expectancy data. The graphs in the simulation can show the overall population counts and the duration of veterans within the model (period of VA obligation to each Veteran).

By 'snapshotting' (using the write signal of the Dataset Writer Block) the Dataset Writer Block, and sub-setting the created entities to those that have not expired, the results are a set of entities and their expected death ages (specifically, their 'created year' and their assigned expected life years), their age at 2030 can be derived and the probability that they will be alive at that age, can be determined from the life tables using conditional probability. This technique counts on the Life Expectancy column of the Life Tables, but demonstrates that you can create or write out data from the simulation to SAS datasets and that those write-outs can be triggered by the model.

A better way to calculate when Entities will exit the model is to use conditional probability. The Life Tables support calculations to answer questions such as "given that you are 35, and will be 49 when the model reaches 2030, what is your probability of death?" By running the simulation well into the future (say to the year 2150), you can capture all the entities added to the model via the Golden Gates block, by stopping the simulation, and saving the entities in that block to a dataset. More generally, once the simulation has stopped, you can save entities from many different blocks to datasets. This gives you a chance to look into the simulation and see the status of the entities at those blocks. Saving entity data to datasets is perhaps the most important feature of SAS Simulation Studio and one that makes this Simulator generally more useful than others.

You can take the data from the Golden Gates Block as a dataset and create a function using procedure FCMP to calculate for each entity age at model entry and the entity age at the year of interest (say entity entered the model in 2018 at age 30 and in 2030, that entity is 42), you can return the probability of that entity being alive at the year of interest. For groups of entities, at each entry age and age at year of interest, you can reduce that count by these conditional probabilities. An example of the FCMP procedure that does this is shown below in Figure 11.

```
%macro getnol;
%global RC;
%let RC=;
%let inreturned=;
proc sql noprint;
  select nol into :inret
    from lt.LifeTable where age = &a ;
  select 1-nol/&inret into :p
    from lt.LifeTable where age = &b ;
quit;
%mend getnol;

proc fcmp outlib = sasuser.ds.functions;
  function getnol_macro(a, b);
    rc = run_macro('getnol', a, b, p);
    if rc eq 0 then return(p);
    else return(.);
  endsub;
run;

data lt.alive2020;
  set lt.goldengates2;
  inc=2020-ClockYear;
  pd=getnol_macro(eage,eage+inc);
  if ClockYear > 2020 then delete;
run;
```

DeathAge	Time	ClockYear	PersonID	Exact Age	Male Life Expentancy	inc	pd
58.27	2070.3	2012	22472	18	58.27	8	0.0107
49	2061	2012	22473	28	49	8	0.0119
45.27	2057.3	2012	22474	32	45.27	8	0.0139
50.86	2062.9	2012	22475	26	50.86	8	0.0114
51.78	2063.8	2012	22476	25	51.78	8	0.0114
43.4	2055.4	2012	22477	34	43.4	8	0.0157
51.78	2063.8	2012	22478	25	51.78	8	0.0114
42.47	2054.5	2012	22479	35	42.47	8	0.0168
57.33	2069.3	2012	22480	19	57.33	8	0.0111
54.54	2066.5	2012	22481	22	54.54	8	0.0115
52.71	2064.7	2012	22482	24	52.71	8	0.0114
49.93	2061.9	2012	22483	27	49.93	8	0.0116
48.07	2060.1	2012	22484	29	48.07	8	0.0122
36.93	2048.9	2012	22485	41	36.93	8	0.0274
50.86	2062.9	2012	22486	26	50.86	8	0.0114
52.71	2064.7	2012	22487	24	52.71	8	0.0114

Figure 11. proc FCMP to create a function call that returns Probability of Death from Life Tables with a small sample (16 rows from a much larger table) of results.

Figure 12 shows the results of the current population aging out by 10 years (many leaving the population in the short term, and the results of adding a steady, but small amount of new Veterans (each with a life expectancy of 50 to 60 years) for 10 years. The result is a distribution still more likely to be in the ages classified as elderly (42% in 2010 and estimated to be 48% in 2020) . The population size is still at about 18 million compared to 10 years earlier (down 3 to 4 million from the 2010 total of about 22.3 million).

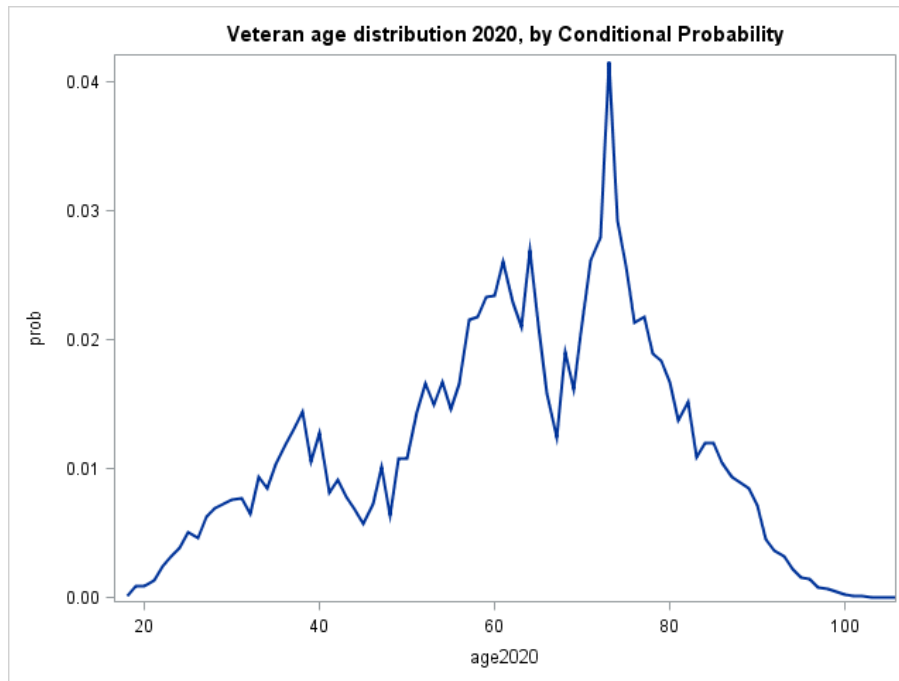


Figure 12. The population distribution of age resulting from the projection to 2020 (about 18 million total).

Figure 13 shows the results of the current population aging out by 20 years (many leaving the population in the short term, and the results of adding a steady amount of new Veterans (each with a life expectancy of 50 to 60 years) for 20 years. The result is a distribution still more likely to be in the older ages with a large percent classified as elderly. However what is not shown is that the population is about ¾ the size of 20 years earlier (at about 14 million). The percent of elderly is projected to be 51% in 2030. Also note that future populations can be more stable (have less dominant peaks or bubbles) as shown in the younger ages of this graph.

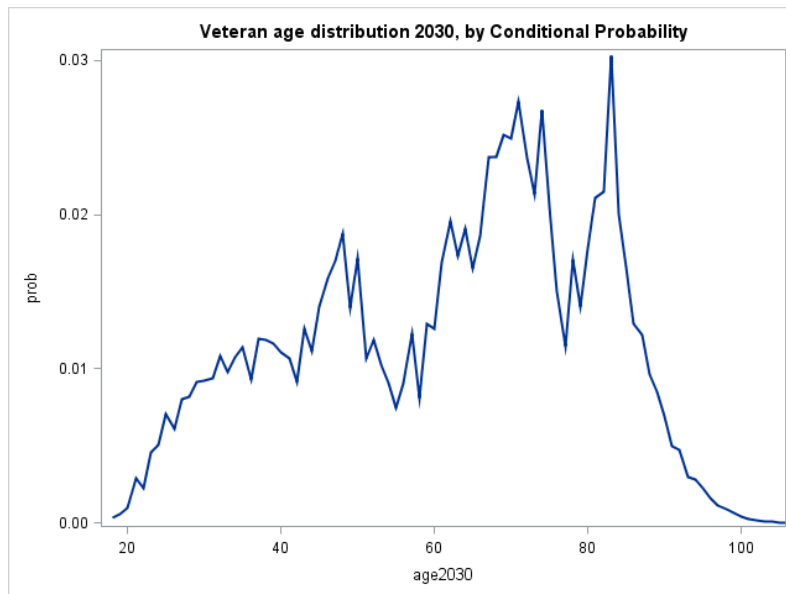


Figure 13. The population distribution of age resulting from the projection to 2030 (about 14 million total)

USE OF AGE DISTRIBUTIONS TO PROJECT SERVICES DEMAND

The use of discrete probability distributions of the age characteristic of a population is useful for planning for service delivery obligations. For example, when the percentage of a population is predominantly over 65, the Social Security Administration has a large service delivery obligation. In this section we will address the service delivery obligations of Compensation and Pension, VA Health Care, and GI Bill benefits. At any time, Veterans may file claims for service connected disability compensation. The 2010 ACS asks about service connected disabilities (yes or no) and asks about the service connected rating (a categorized percent)⁷. The ACS (and the past decennial surveys) shows that Veterans get disability compensation at the ages shown in the discrete probability distributions below. Figure 15 shows the discrete age probability distribution for the Veterans who answered yes to the question as described above and you may notice the 'spike' at the same place as the overall Veteran population in Figure 3. To disambiguate the population 'spike' from the 'spike' of Veterans getting a disability, Figure 16 shows both distributions on a common graph and adds a difference line (diff).

⁷ Question 28 of the 2010 ACS questionnaire can be found here, <http://www.census.gov/acs/www/Downloads/questionnaires/2010/Quest10.pdf>

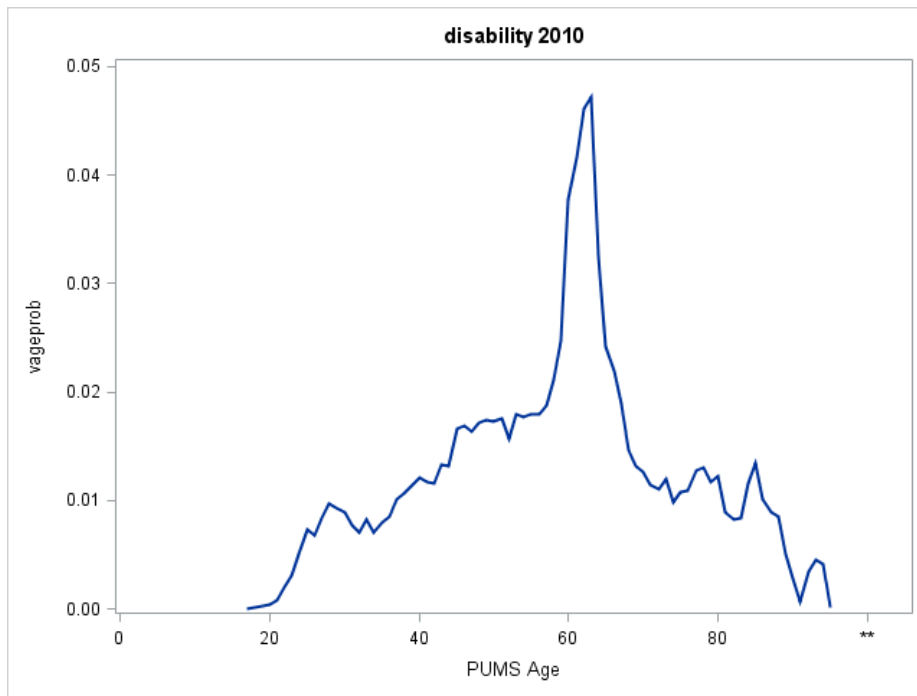


Figure 14. The 2010 Veteran population distribution of age by answer to question 28 a. “Does this person have a service-connect VA disability rating?”

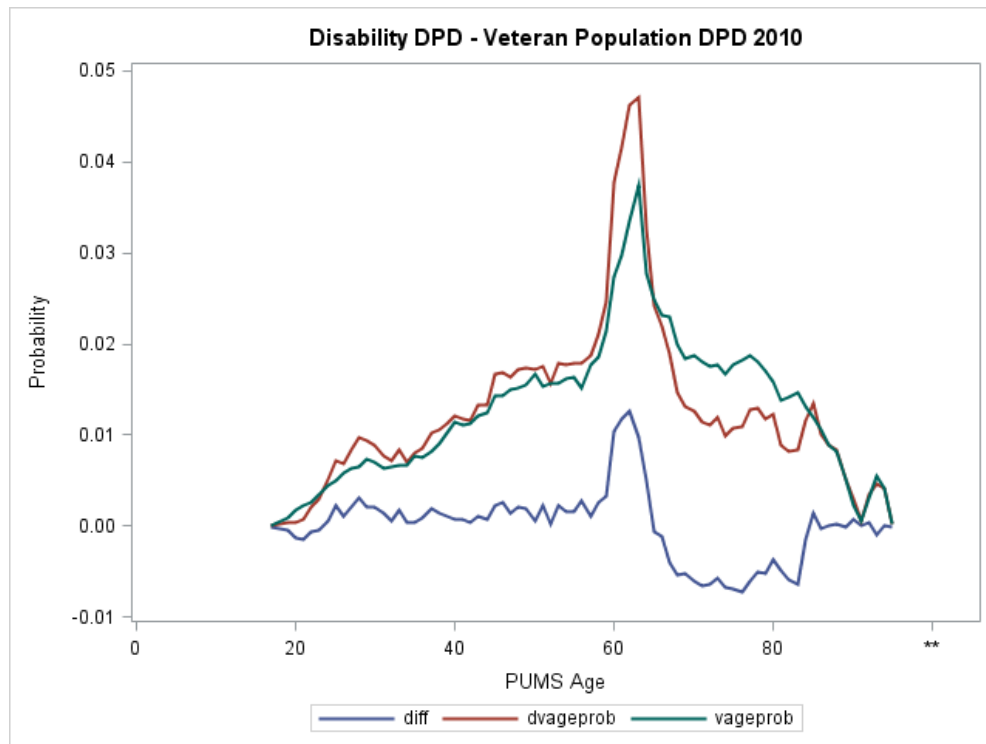


Figure 15. Figure 13 shown relative to the overall Veteran Population distribution (Figure 3) with a difference line.

This difference line shows an increase in use of the compensation and disability benefit at the 60 to 65 age category. Also notice that the benefit is not used in proportion to the population after that age category (at least on this graph). The diff line in Figure 16 explains why the Veteran Benefit Administration is seeing a dramatic increase in claims activity starting in 2009 as reported by the Monday Morning Workload Reports⁸ and graphed below in Figure 16 using proc SGPLOT and sequentially imported Excel spreadsheets.

⁸ This data made available to the public at <http://www.vba.va.gov/REPORTS/mmwr/index.asp>

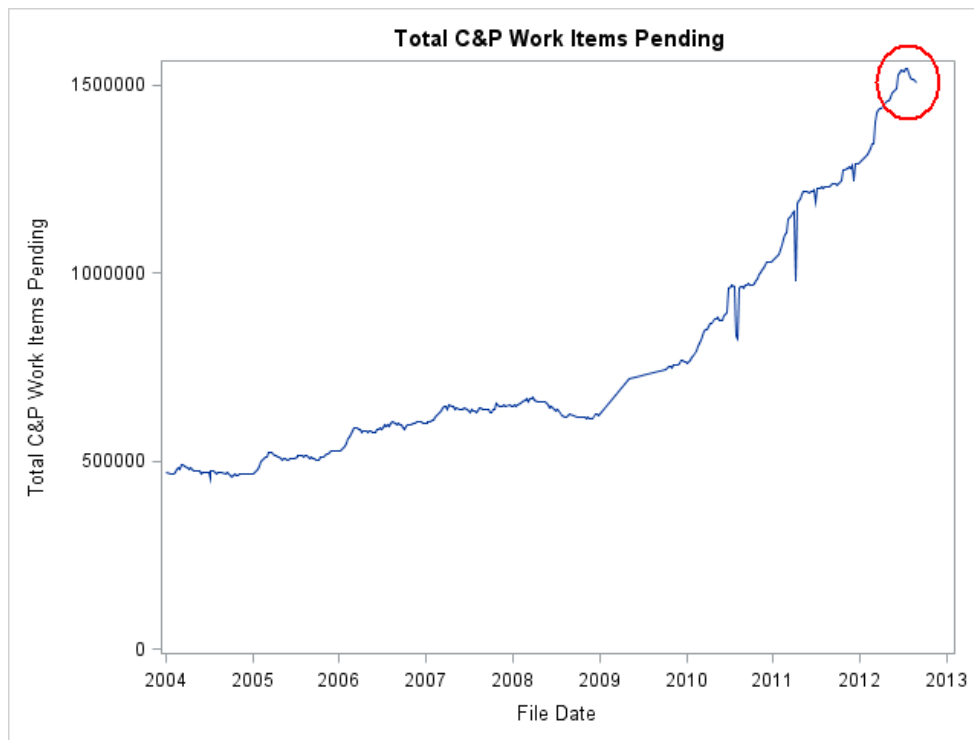


Figure 16. Data from the VBA’s Monday Morning Workload Report on Claims pending.

Examining Figure 17 may lead to a better understanding of why the Claims activity shown in Figure 16 is occurring. As the Veteran’s population peak moves over the static utilization peak, claims activity will also peak (as they sum), but then after the population peak passes over the utilization peak, the claims activity should recede and Figure 16 should change direction (and it has, as shown circled in red on Figure 16). Closer examination (of the 1990 Veteran population for example) reveals that the utilization peak and troughs are generally due to VA compensation and disability rules (passed by Congress) that are conditional upon age 65. The fact that Veterans retire from their civilian work lives at age 65 may also play a part.

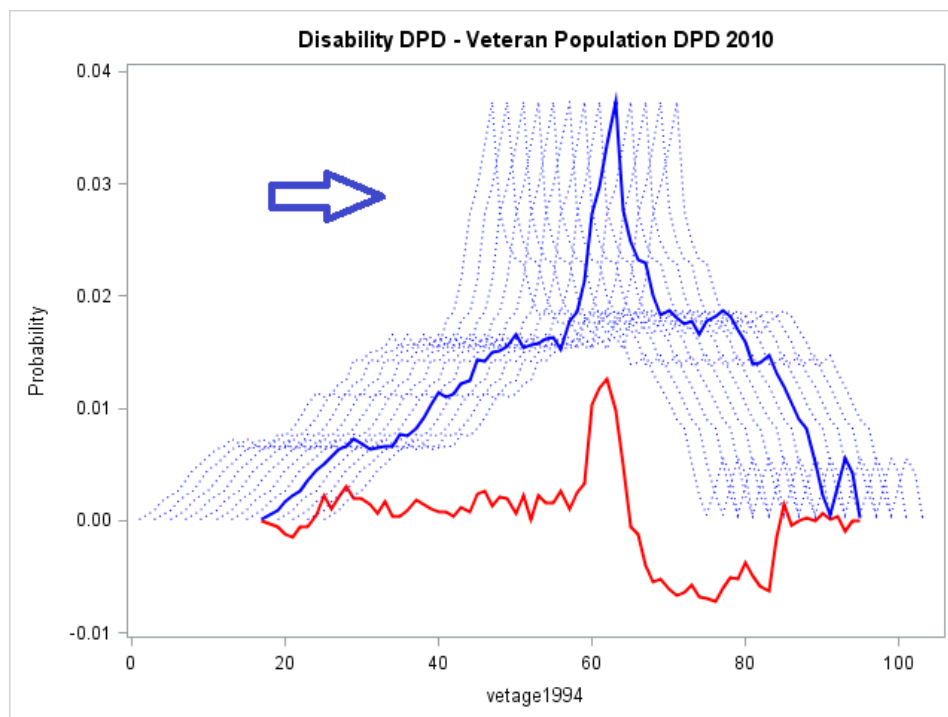


Figure 17. Trying to show the population spike moving in time past the utilization spike.

The use of VHA hospital care can be addressed with the yes answers 2010 ACS question 16, which asks if “Is this person CURRENTLY covered by any of the following types of health insurance or health coverage plans?” and answer f. is “VA (including those who have ever used or enrolled for VA health care)”. This question is designed to show health care utilization for a wide variety of health care providers and should be a useful question for many health care related studies. The resulting relative distribution graph is shown in Figure 18 below.

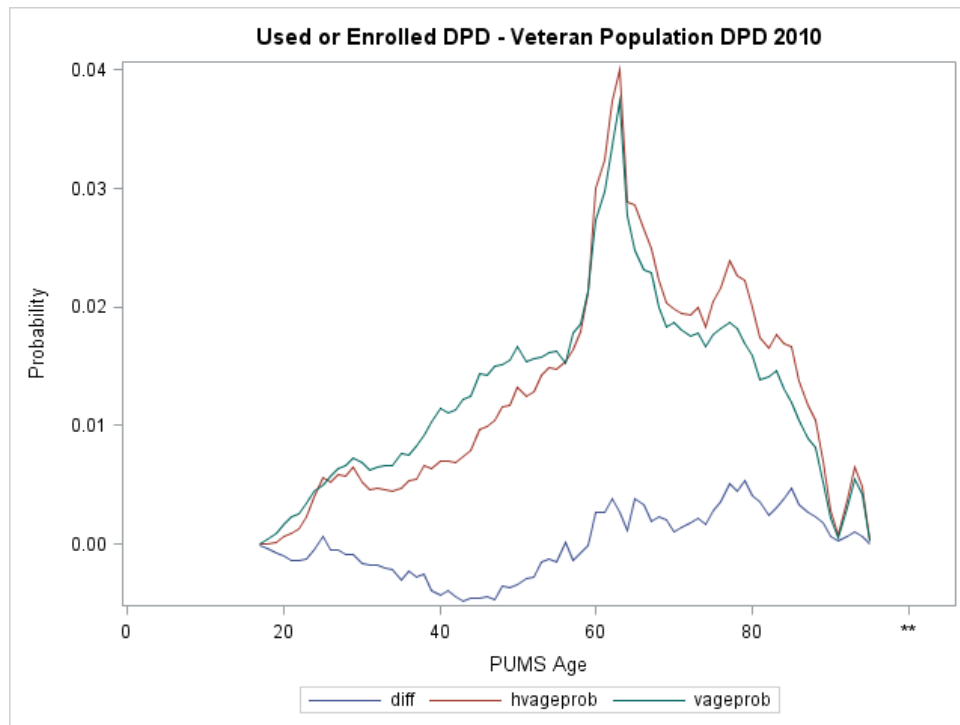


Figure 18. VA health care use shown relative to the overall Veteran Population distribution (Figure 3) with a difference line.

VA health care use seems to track fairly well with the population but shows a preference for less use prior to age 65 and more use after that age. If the Veteran population is predominantly over 65 this service delivery obligation is consequently increased.

The use of the GI Bill service delivery can be estimated using data from a publicly available data handout used⁹ at an American Council of Education seminar in 2008. This data supports a histogram shown in figure 19 below called “K. Wilson’s Edu Percent”. The derived blue line is basically a shift of the new Veterans data in Figure 4 to the right by about 3 or 4 years and adding a bump for the 43 to 47 age group. The conjectured reason for this ‘bump’ is that GI Bill benefits run out after 10 to 15 years after becoming a Veteran and a small group of GI Bill benefit users wait till near the end of their eligibility. This is perhaps the best example of an age dependent service delivery.

⁹ Presentation and Handouts from the American Council on Education seminar titled “Serving those who server: Veteran’s Summit (2008)”
http://www.acenet.edu/Content/NavigationMenu/ProgramsServices/MilitaryPrograms/Vet_Summit_June_08_.htm,
http://www.acenet.edu/Content/NavigationMenu/ProgramsServices/MilitaryPrograms/Wilson_Keith.pdf,
<http://www.acenet.edu/Content/NavigationMenu/ProgramsServices/MilitaryPrograms/WilsonDemographicData.pdf>

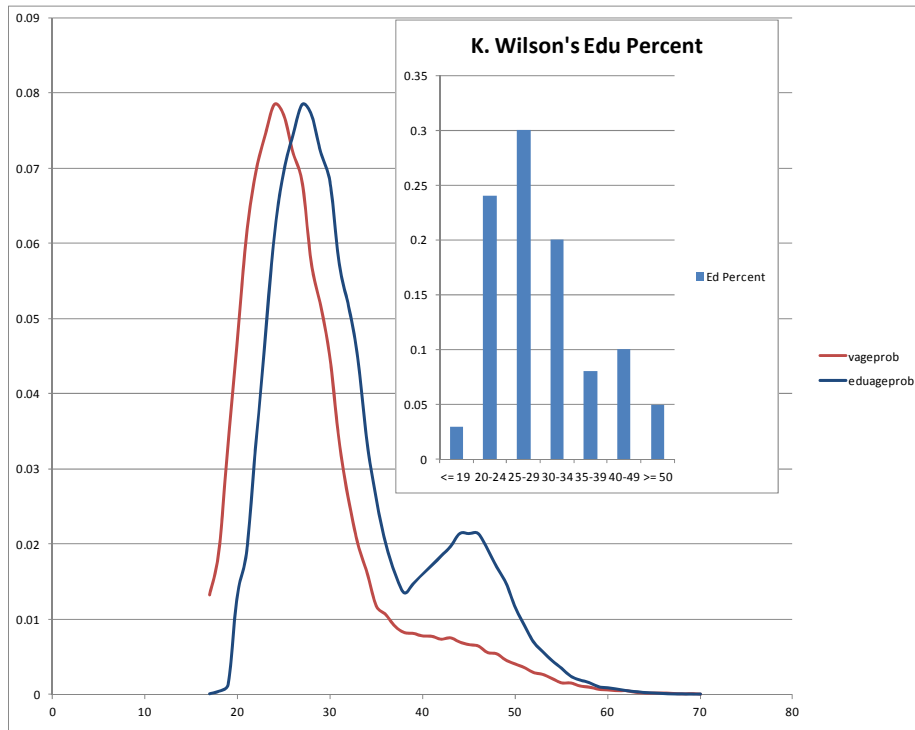


Figure 19. A hypothetical Education Benefit use distribution shown relative to the ‘New’ Veteran Population distribution (Figure 4).

ACKNOWLEDGEMENTS

This material is based upon work supported in part by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development. Access to the remarkable VINCI facility is provided by the Office of Research and Development and is especially appreciated. More generally, I am grateful to the Department of Veterans Affairs for the use of their SAS software resources and facilities without which this paper would not have been possible.

I am also grateful to the Bureau of the Census for their efforts on the American Community Surveys, the 1990 and 2000 decennial surveys. All of the data used in this paper, primarily the 2010 American Community Survey and other Census products, is publicly available and should be referenced.

Finally, and again, the views expressed in this paper are those of the author and do not reflect the official policy or position of the Department of the Veterans Affairs, or the U.S. Government.

CONCLUSION

This paper described using the survey procedures in SAS and a stratified survey to glean current population characteristics in support of projecting that population forward in time with SAS Simulation Studio®. The ability to go from complex statistical procedures to a sophisticated Operations Research technique represents a chain of capabilities unmatched in any other environment.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Mike Grierson
 mcg@mcg-ct.com
www.mcg-ct.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration. Other brand and product names are trademarks of their respective companies.