PVS Membership Study: Status Update

Jim Thompson
PVS Strategic Planning Chair

FOR BOARD DISCUSSION PURPOSES ONLY

Purpose

- Background: Recommendation from the November 14th BOD meeting:
  - Tell the story of how PVS serves the needs of the community
  - Provides the Executive Director and Board members information to engage key stakeholders of aquatic facilities and government bodies
    - Document PVS needs for aquatic facilities
    - Document benefits (impacts) of meeting (failing to meet) PVS needs

- Today's discussion provides status of the work to analyze PVS membership
  - Share results of the analyses with BOD

- Looking BOD members to answer
  - What questions do you have on the analyses?
  - What directions would you like to see the analysis work go?

Today's Discussion

- County of Residence
- Membership Forecast
- Other Youth Sports Programs
- Wiki Collaboration Site
- Key Points
- Next Steps
- Appendix
  - Geocoding
  - Forecast Regression Models

County Of Residence Determination

- Two methods available to determine “county of residence”
  - Translating Zip Code to county using publically available data. This method was utilized for the December BOD presentation.
  - Geocoding individual addresses to determine Federal Information Processing Standard (FIPS) county designation. FIPS designations are used in governmental analyses.
- Zip Code method used in initial analysis presented at the December 2007 BOD meeting.
  - Zip codes do not uniquely align with counties. Some zip codes overlap multiple counties.
- December 2007 BOD meeting authorized spending up to $500 to support this demographic study.
  - Subscribed to a geocoding service (www.geocode.com) to perform geocoding of individual addresses

County Of Residence Determination

- Geocode 2007 PVS Registration data as of 11/15/2007. (See Appendix for detail results)

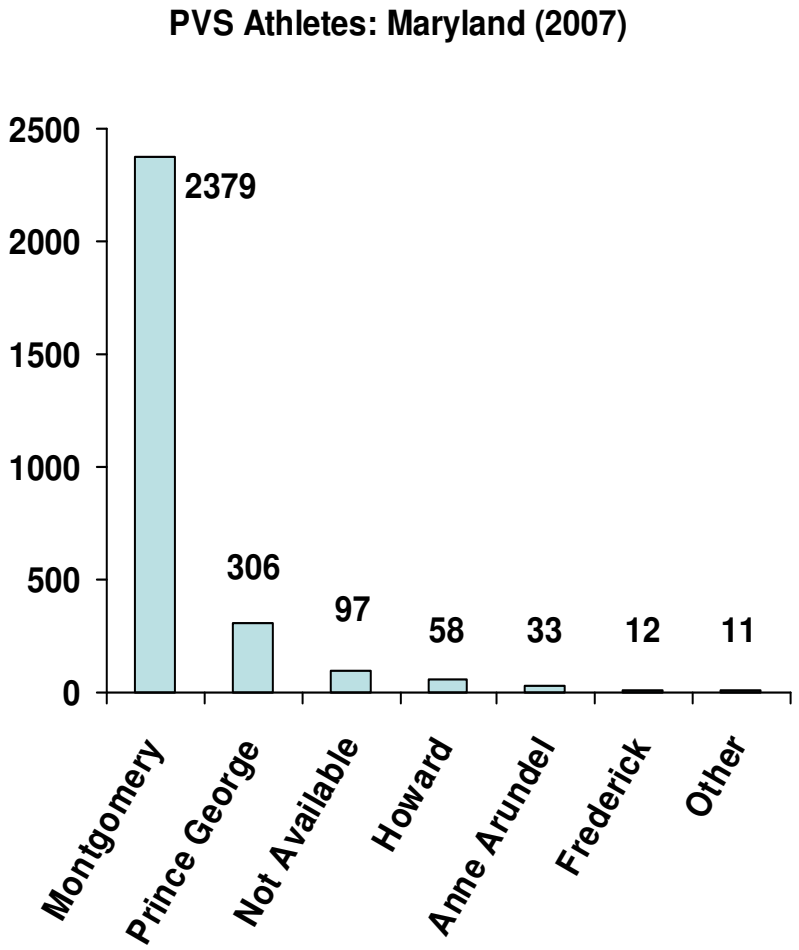
8,217 Registered Members in 2007
-93 Members with mailing addresses outside of DC, MD, VA
-32 Addresses with obvious errors or P.O. Boxes
8,092 Addresses to geocode

Reduced above by eliminating duplicate addresses, e.g., siblings at the same address
5,862 Sent to service for geocoding

Geocoding service returned
5,686 Addresses were successfully geocoded
176 Addressess encountered problems in geocoding

Resulting in
7,855 Members with geocoded addresses
237 Members with addresses that encountered problems
125 Members whose addresses were not sent for geocoding
8,217

PVS Membership Distribution (Geocoding)



Maryland

- Membership is concentrated in Montgomery County, which has about 82% of Maryland athletes (29% overall)
- Maryland has 2,896 athletes (35% overall)

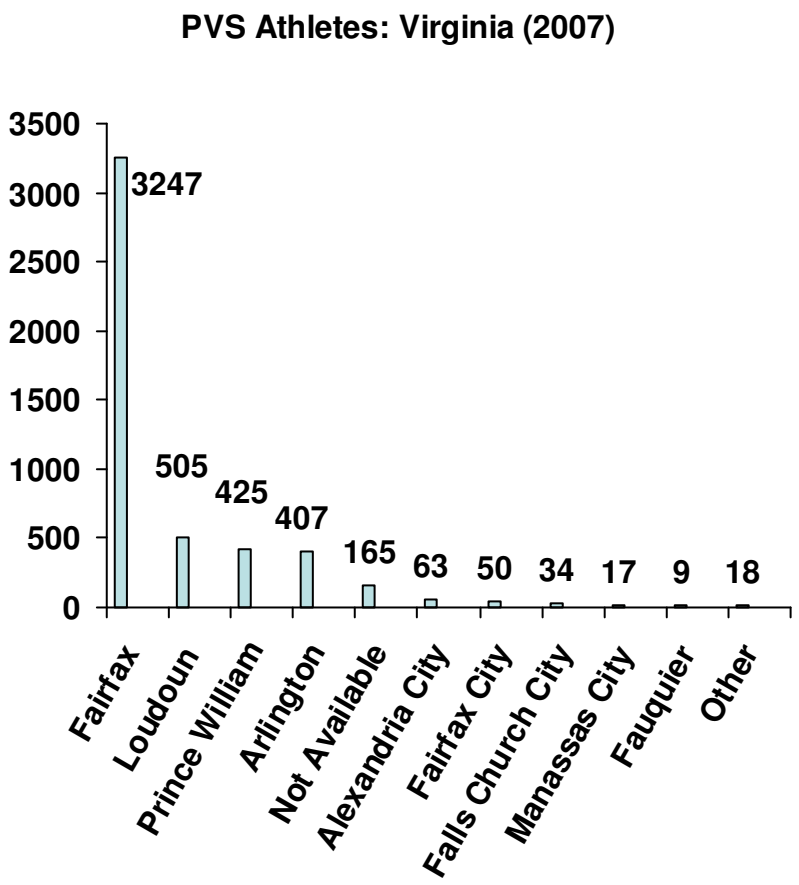
District of Columbia (2007)

- DC has 288 athletes (3.5% overall)

Notes

- Others: Mailing addresses outside of DC, MD and VA account for 93 athletes (just over 1% overall)
- Not Available: Not sent for geocoding or experienced error during geocoding

PVS Membership Distribution (Geocoding)



Virginia

- Like Maryland, the bulk of PVS members is concentrated in one county (Fairfax), which accounts for 66% of Virginia members (40% overall)
- Virginia has 4,940 athletes (60% overall)
- Item of note: With the creation of Patriot Swim Club (PATS) about 150 swimmers will move to VA LSC.
  - 122 from Prince William County
  - 15 from Fairfax County
  - 9 unknown county
  - 3 from Fauquier County
  - 2 from Loudoun County
  - 2 from Manassas City

Chart at left includes PATS swimmers

County Of Residence Determination

Results Summary

- Compared to Zip Code method there was little difference for DC or MD addresses with geocoding
- However, for VA addresses geocoding led to significantly different results in some areas:
  - Fairfax City: 174 (zip code) to 50 (geocoding)
  - Fairfax County: 3,193 (zip code) to 3,247 (geocoding)
  - Falls Church City: 57 (zip code) to 34 (geocoding)
  - Manassas City Park: 103 (zip code) to 4 (geocoding)
  - Prince William County: 310 (zip code) to 425 (geocoding)
- Effects of this change
  - Market penetration percentages change
    - Fairfax City from 4.78% to 1.37%
    - Falls Church City from 3.20% to 1.91%

County Of Residence Determination

- Effects of the change (continued): The top five counties in VA for PVS membership in 2007:

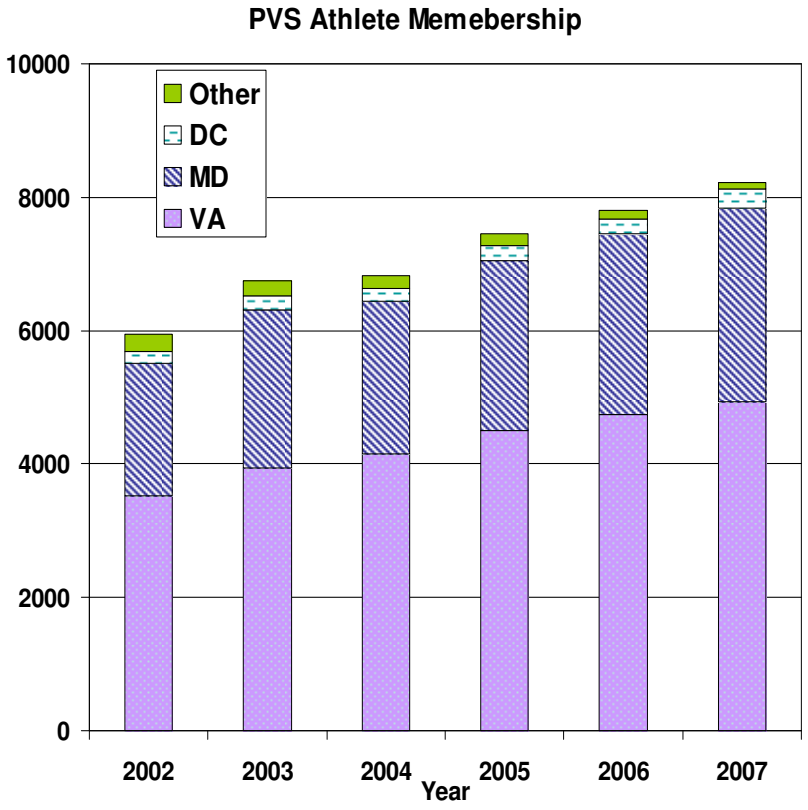
Zip Code Method	Geocoding Method *
Fairfax (3,193)	Fairfax (3,247)
Loudoun (538)	Loudoun (505)
Arlington (410)	Prince William (425)
Prince William (310)	Arlington (407)
Fairfax City (174)	Alexandria City (63)

* 149 addresses were not geocode or experienced geocoding errors.

Zip Code vs Geocoding

- Recommendation: Use geocoding method to determine county of residence
  - Geocoding provides FIPS county designation. FIPS is a government standard for designating geographic information and is used by other entities (e.g., government & business)
  - Using FIPS designation allows for “Apples to Apples” comparison when incorporating Census data.
  - Geocoding provides census tract results if finer level of granularity is needed
  - Some zip codes do not uniquely identify a county
- Future Implications:
  - Effort to scrub addresses to reduce geocoding errors
  - Ongoing cost to geocode new addresses
    - Brute-force geocode all addresses, instead of eliminating duplicates, to reduce possibility of errors in data manipulation
- In the remainder of this presentation, unless noted otherwise, the geocoding method is the basis for any results showing county names.

Forecasting PVS Membership



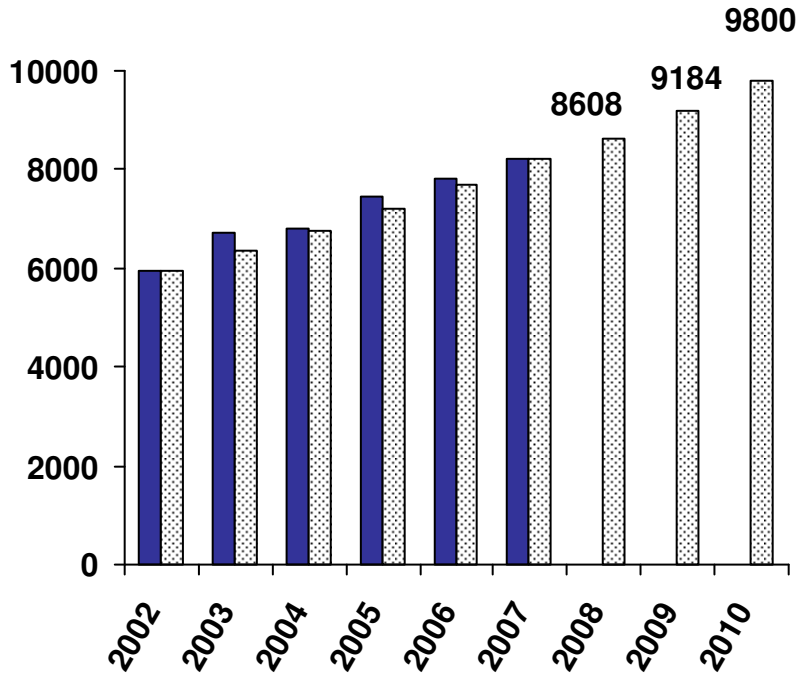
- "Other" category represents PVS athletes with mailing addresses outside of DC, MD or VA.
  - Seasonal Athletes or
  - Athletes who have subsequently moved outside of PVS but have maintained USA Swimming registration. Registration data only pulls current address of athlete.
- Implies about a 6.7% compounded annual growth rate (CAGR) from 2002 thru 2007.

PVS Membership Forecast I

Preliminary

Three-year Forecast

PVS Athlete Membership*



*** Past performance does not guarantee future results.**

Compounded Annual Growth Rate (CAGR)

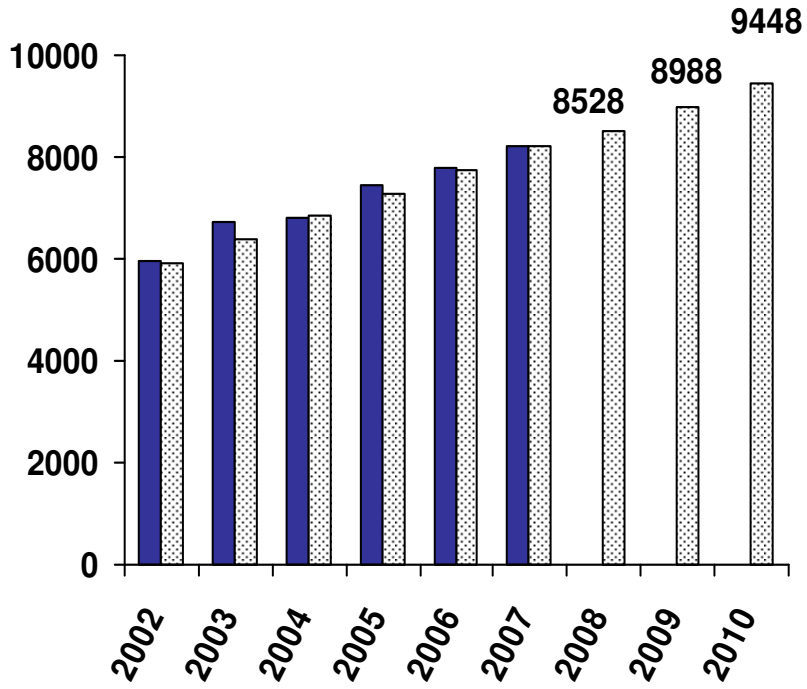
- Assuming 6.7% CAGR, PVS could expect to have membership in the upper-9,000s by 2010. Includes adjustment for departing PATS swimmers.
- Solid bars are actual PVS membership numbers and patterned bars are forecasts based on the model.
- Where actual data exists, average relative error is 1.87%

PVS Membership Forecast II

Preliminary

Three-year Forecast

PVS Athlete Membership*



*** Past performance does not guarantee future results.**

Census Data Based Regression Model

- Regression Model:
  - Dependent Variable: PVS membership
  - Explanatory Variable: Population (5 to 19 year-olds) and geography (aggregation of selected counties).
- Developed specific regression models for VA and MD. Assuming DC is a small but constant factor at this time.
  - Both regression models appear robust with population data provided by respective State planning departments.
- By 2010 membership could exceed 9400 members.
- See Appendix for details
  - Includes adjustment for departing PATS swimmers
- Average relative error is 1.45%.

Modelling Commentary

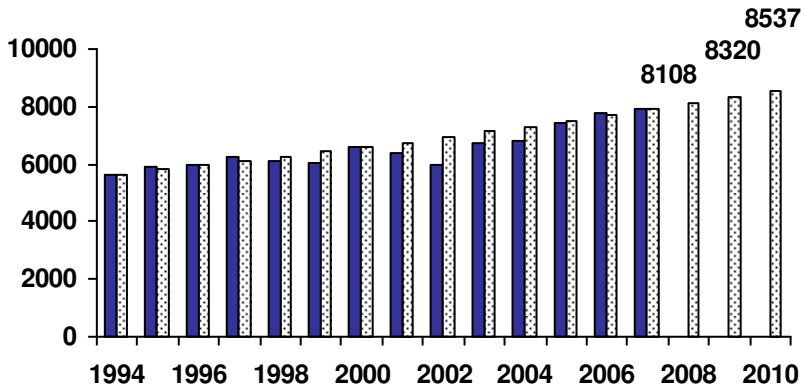
Three Views of PVS Membership History			
Year	**Data Set 1**	**Data Set 2**	**Data Set 3**
1994	N/A	5653	N/A
1995	N/A	5877	N/A
1996	N/A	5998	N/A
1997	N/A	6249	N/A
1998	N/A	6132	N/A
1999	N/A	6038	N/A
2000	N/A	6625	N/A
2001	N/A	6391	N/A
2002	5941	5945	N/A
2003	6739	6741	6722
2004	6821	6823	6806
2005	7452	7453	7421
2006	7803	7804	7797
2007	8217	7902	8188

- All models are based on assumptions. For the forecast models in this presentation the history of PVS membership is one of the assumptions. The table on the left show three views of this history.
- Data Sources:
  - "Data Set 1" – provided by J. Garner to bootstrap the demographics analysis. This set used in the forecast models described in this presentation.
  - "Data Set 2" – provided by J. Hirschmann as historical data
  - "Data Set 3" – extracted from PVS Registrar's reports (dated September of the respective year)

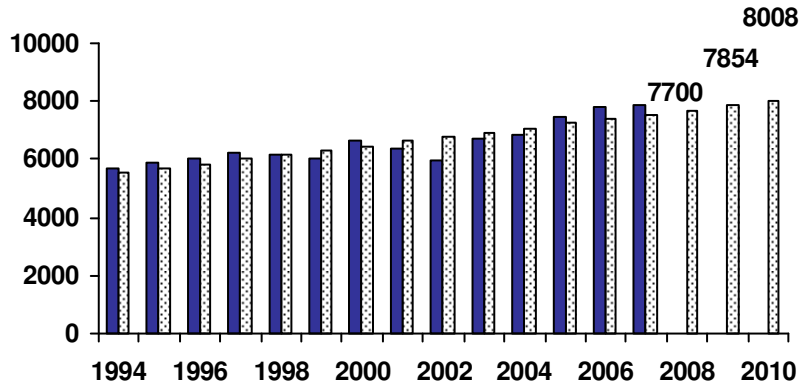
Modelling Commentary

Preliminary

Data Set 2 - CAGR Forecasts



Data Set 2 - Regression Forecast



- Different assumptions can lead to different conclusions. The charts on the left show PVS membership forecasts based on the Data Set 2 historical data.
- When compared to Forecasts I & II, there is a forecast difference of up to 1,800 members in 2010.
- A review procedure is required to ensure correct assumptions are feeding into the analyses. One possible recommendation:
  - Administrative Vice Chair, or delegate, reviews and approves demographic assumptions and analyses.
  - Operations Vice Chair, or delegate, reviews and approves meet-level assumptions and analyses.
- One criterion for selecting a set of assumptions:
  - consistency with budget planning

PVS Membership Forecast Summary

- Both forecast models show similar results
  - Four to five percent growth in 2008
  - Possible for membership to reach mid to upper 9,000s by 2010
  - Factors affecting membership that are not explicitly modelled:
    - Local economic and demographic (CAGR only) conditions
    - Swim club limitations, i.e., pool lane rentals
    - Olympic Year effect
- CAGR Model
  - Pro: Simple, easily calculated
  - Con: Does not account for in local factors
- Regression Model
  - Pro: Accounts for some local factors (geography & demographics)
  - Con: More complex to use, calculation or usage errors may go unnoticed.
- Preferred Model: Regression Model
  - Regression model forecasts utilize data from government sources, e.g., Commonwealth of Virginia and Maryland's Department of Planning
  - Regression model's foundation ties to local conditions, i.e., local demographics.
  - Use CAGR as a secondary model for validation purposes (sanity check)

PVS Membership Forecast Summary

- Both CAGR and Regression Models
  - Require periodic updates to model factors to reflect changes in the environment and actual results.
- Future work
  - Investigate if further stratification of data will improve forecast, e.g., separate regression models by gender and individual counties
  - Investigate other explanatory variables, such number of households, number of members in a household, median income.
- Uses for the forecast
  - Since membership fees are a significant portion of income, factor membership forecasts into longer term budget planning process
  - Share forecasts with regional recreational entities and local government officials

Other Youth Sports Programs

- Guidance from December Board meeting:
  - "The value of obtaining the number of individuals involved in other sports (i.e., soccer, lacrosse, etc.) so as to enable comparisons with swimming"
- After several days of Internet search unable to find a central clearinghouse for youth sports programs in the Washington, D.C. area.
- Focusing on Fairfax County and using publically available information on the Internet found potential information sources. Examples:
  - Chantilly Youth, Inc (http://www.chantillyyouth.org) claims 14,000 participants covering 11 sports. This implies about 1,200 participants per sport.
  - Braddock Road Youth Club (http://www.bryc.org/) cites 3,000 participants for 7 sports. This implies 430 athletes per sport.
  - Unclear how these sources inter-relate, potential to double count, for example, Fairfax County Youth Football League cites Chantilly Youth, Inc. and Braddock Road Youth Club as members.
- Suspend current work in this area and focus on other types of analyses
  - Not clear how to proceed in this task
  - Assistance with this task is needed

Wiki Collaboration Site

- Established a private wiki to support distributed collaboration for the demographic analysis.
  - Free wiki service (www.wetpaint.com); some limitations
- Private wiki is viewable only to participants who are invited to join
  - At this time, four Board members who expressed an interest the work, Web Master and Registrar are able to access
  - Other Board member who wish to participant or oversee the work should request an invitation
- Wiki currently used for
  - Repository of ideas
  - Posting raw analyses for review, e.g., since late December the following have been on the wiki for review
    - Membership forecast models
    - Hy-tek meet analyses
  - Preview of draft presentations
  - On-line discussion
  - Content refinement

Wiki Collaboration Site

- Wiki capabilities
  - Basic WYSIWYG editing
  - Maintain history of page versions
  - Email notification when content is updated
  - Threaded discussion forum
  - Keyword search within wiki
- Other purpose for the wiki – Demonstrate how wiki technology could support on-line collaboration of geographically dispersed participants. A possible tool to facilitate the Board's work.

Key Points

- Geocoding is the preferred method for incorporating county level data
  - Especially in results shared with external entities
- Membership forecast
  - Four to five percent growth in 2008
  - Membership growing into the mid to upper-9,000s by 2010
  - Not all influences on membership can be explicitly modelled
  - Factor membership forecasts into longer term budget planning process
  - Craft message to regional recreational entities and local government officials on the outlook for PVS
  - Adjustment to models to reflect actual experiences and changes in the environment
- Review process required for analyses
- Youth Sports
  - Suspending work in this area, focus efforts on other analyses
- Wiki Technology
  - Supporting the collaboration on the analysis work
  - Possible tool for supporting other PVS work

Next Steps

- Establish review procedure for forecast models and analyses.
- Future analyses
  - February BOD Meeting: Hy-tek Meet Manager data
    - What are the demographics of meet participants?
    - How are meets geographically distributed?
    - What trends are developing?
  - March BOD Meeting: PVS Demographics
    - Finer granularity analysis on PVS membership
- What directions would the Board like to see the analysis work go? Possible research topics:
  - Refine PVS membership forecast model. Some possible extensions:
    - individual county-level
    - include other explanatory variables such as income levels, household composition and gender.
  - Model meet participation levels based on PVS membership
  - Incorporate modeling aspects into the budget planning process

Appendix

Geocoding

Comparison of Geocoding vs Zip Code Methods For County Determination

	Geocoding	Zip Code
DC		
District of Columbia	281	288
Geocoding Errors	7	
Total	288	288
MD		
Anne Arundel County	33	34
Baltimore County	1	1
Calvert County	8	9
Charles County	1	1
Frederick County	12	12
Harford County	1	1
Howard County	58	59
Montgomery County	2379	2448
Prince George's County	306	325
Geocoding Errors/Unknown	81	6
Not sent for geocoding	16	
Total	2896	2896

	Geocoding	Zip Code
VA		
Alexandria city	63	76
Arlington County	407	410
Fairfax city	50	174
Fairfax County	3247	3193
Falls Church city	34	57
Fauquier County	9	9
Frederick County	3	3
Loudoun County	505	538
Manassas city	17	52
Manassas Park city	4	103
Prince William County	425	310
Stafford County	4	2
Other	7	7
Geocing Errors/Unknown	149	6
Not sent for geocoding	16	
Total	4940	4940

Geocoding

Geocoding Errors Encountered

Amount	Description
2	Invalid locality name, e.g., City contained “Derwood20855”
131	Street name could not be found in requested locality, e.g., “Pouring Rain Pl” in Nokesville, VA or “Calvert Hall” in College Park
95	Address range for input house number did not exist on given street, e.g., 12610 Bayard Drive, Reston, VA
8	More than one segment with adequate address range, e.g., “12610 Bayard, Reston, VA”, possible “Bayard St” or “Bayard Dr” or “Bayard Ave”
1	Caused by programming error

Regression Model

Regression model: $EstPVSXX = b \times age5\_19 + c$, where *EstPVSXX* is the estimated number of PVS athletes in Maryland or Virginia (*XX*=*MD* or *VA*, respectively) and *age5_19* is the Census estimate for the population aged 5 through 19 in specific geographic areas in the respective states.

For MD, the counties of Anne Arundel, Howard, Montgomery and Prince Georges were selected for modelling. The following geographic areas were used in the Virginia model: the cities of Alexandria, Fairfax, Falls Church and Manassas; and the counties of Arlington, Fairfax, Loudoun and Prince William. These areas were selected because they contained a large portion of PVS membership

The following are the raw data for modeling:

PVS Membership		
Year	PVSMD	PVSVA
2002	1989	3526
2003	2385	3935
2004	2303	4144
2005	2539	4508
2006	2703	4750

Population	Ages 5 thru 19	
Year	VA	MD
1990	282,590	
2000	365,633	517,682
2005		553,068
2010	439,752	564,659
2015		557,229
2020	484,544	561,845
2025		568,736
2030	534,879	585,605

Maryland: 2007 Total Population Projections by Age, Race and Sex (8/2/07), Maryland Department of Planning, Planning Data Services.
Virginia: http://www.vec.virginia.gov/vecportal/lbrmkt/popproj.cfm, pulled 12/30/2007, Virginia Employment Commission.

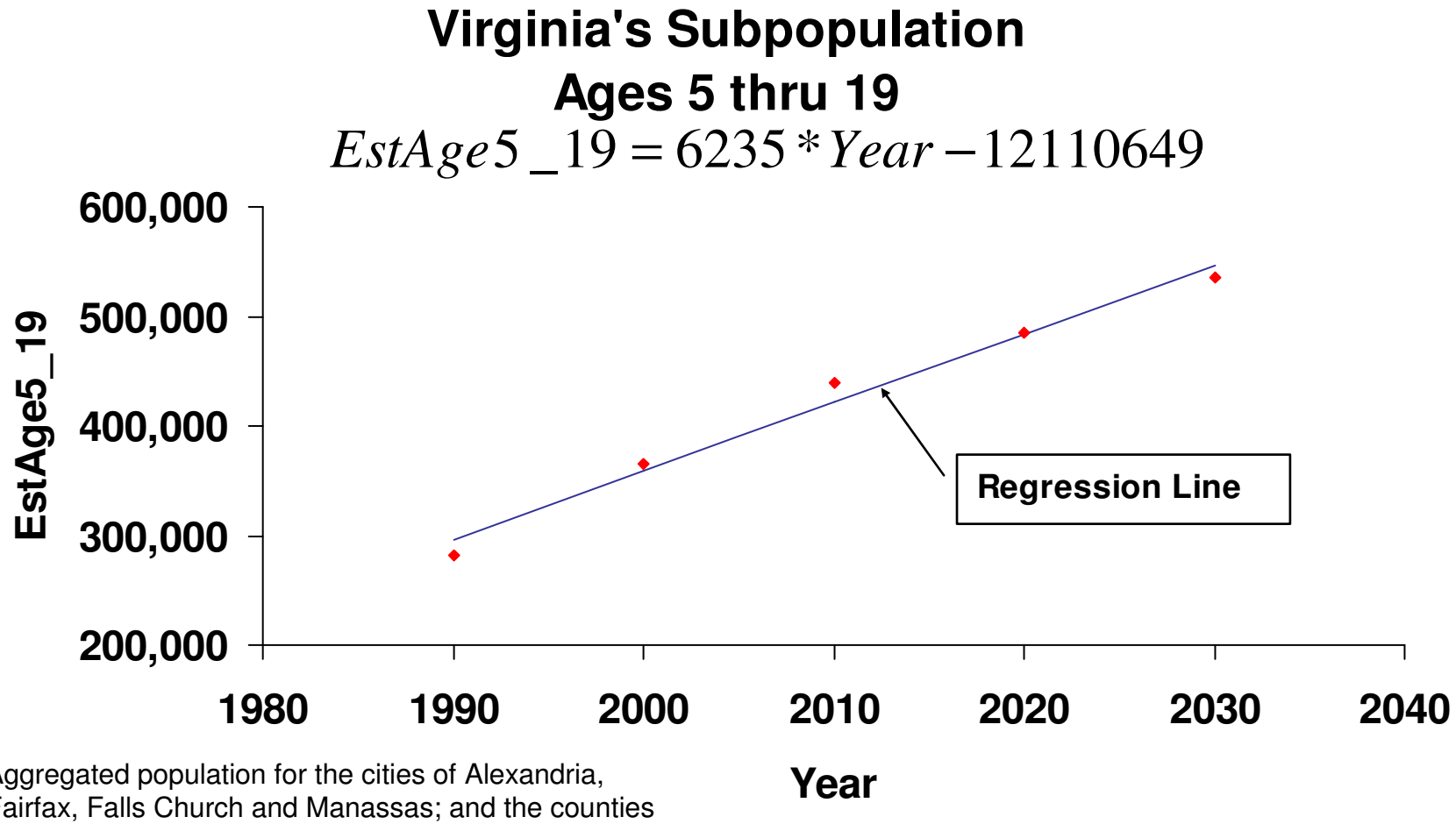
Regression Model

First, a regression model was fitted to MD and VA population data to forecast the number of persons in the 5 to 19 year-old subpopulation. The forecasted variable is denoted by *Estage5_19*. The PVS data is then merged with the subpopulation forecast for the years with PVS data. This merged data set is used to develop the regression models for *EstPVSDC* and *EstPVSMD* (see table).

	Virginia		Maryland	
Year	**EstAge5_19**	**PVSVA**	**EstAge5_19**	**PVSMD**
2002	371,600	3,526	536,834	1,989
2003	377,835	3,935	538,493	2,385
2004	384,070	4,144	540,152	2,303
2005	390,305	4,508	541,811	2,539
2006	396,540	4,750	543,470	2,703

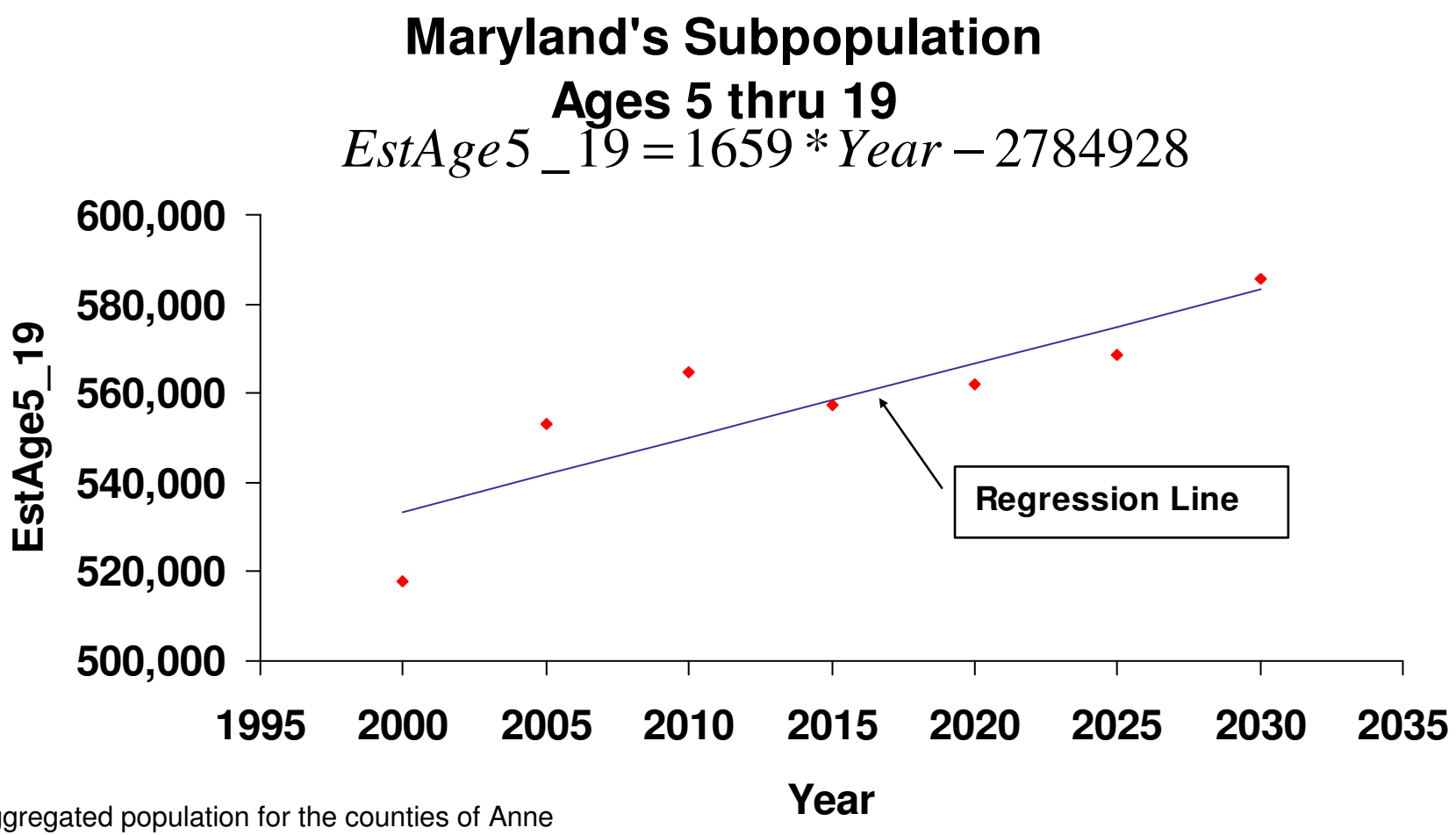
Depending on time period, one of two functional forms are used to estimate overall PVS membership (*EstPVSALL*). For the period 2002-2007, *EstPVSDC* and *EstPVSMD* are summed with a constant factor representing DC athletes (*EstPVSDC*). For this analysis *EstPVSDC* is 280. For the period 2008-2010, *EstPVSVA*, *EstPVSMD* and *EstPVSDC* are summed. From this sum the number of athletes in the newly created PATS swim club is subtracted.

Regression Model



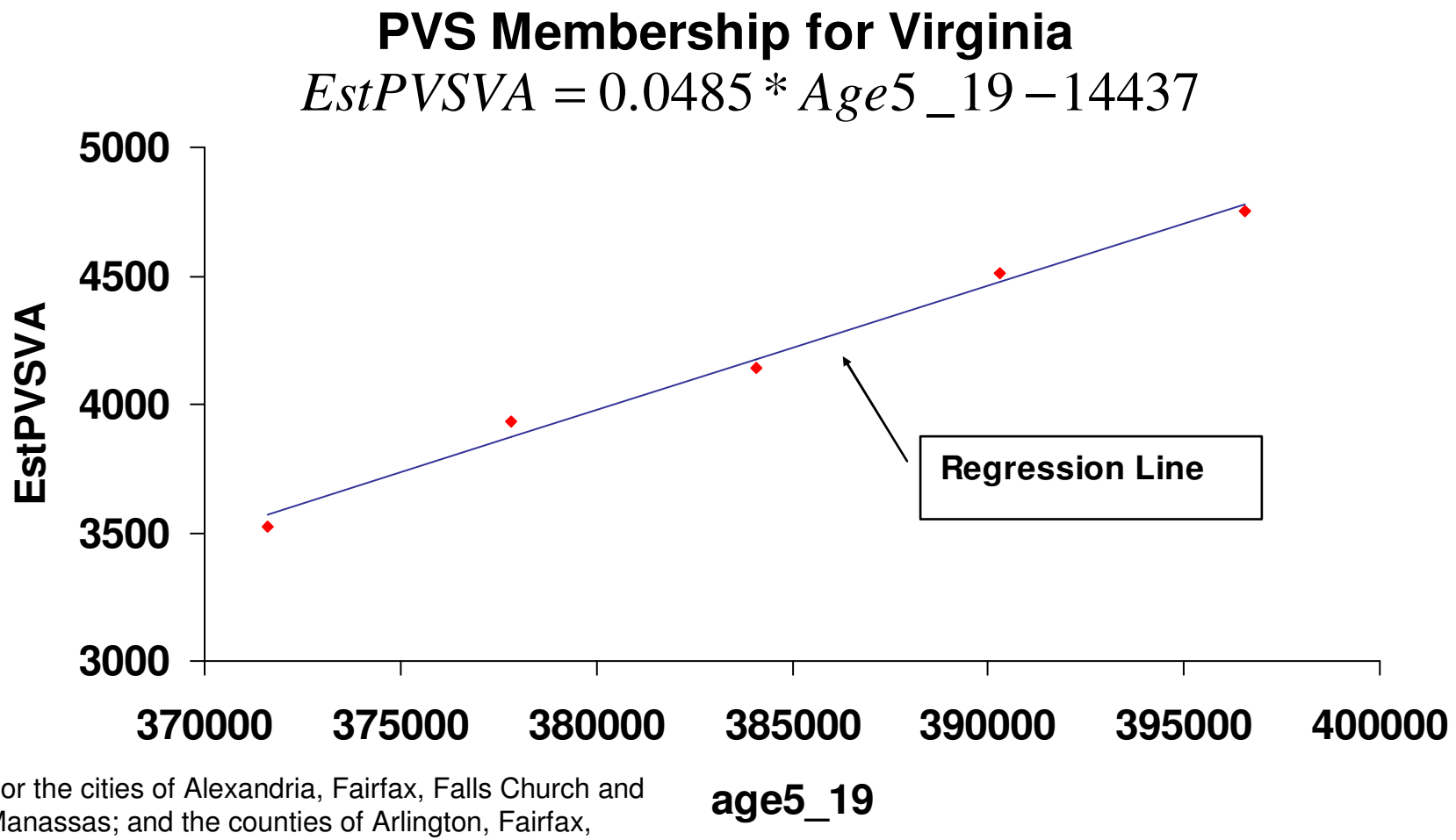
Aggregated population for the cities of Alexandria, Fairfax, Falls Church and Manassas; and the counties of Arlington, Fairfax, Loudoun and Prince William.

Regression Model



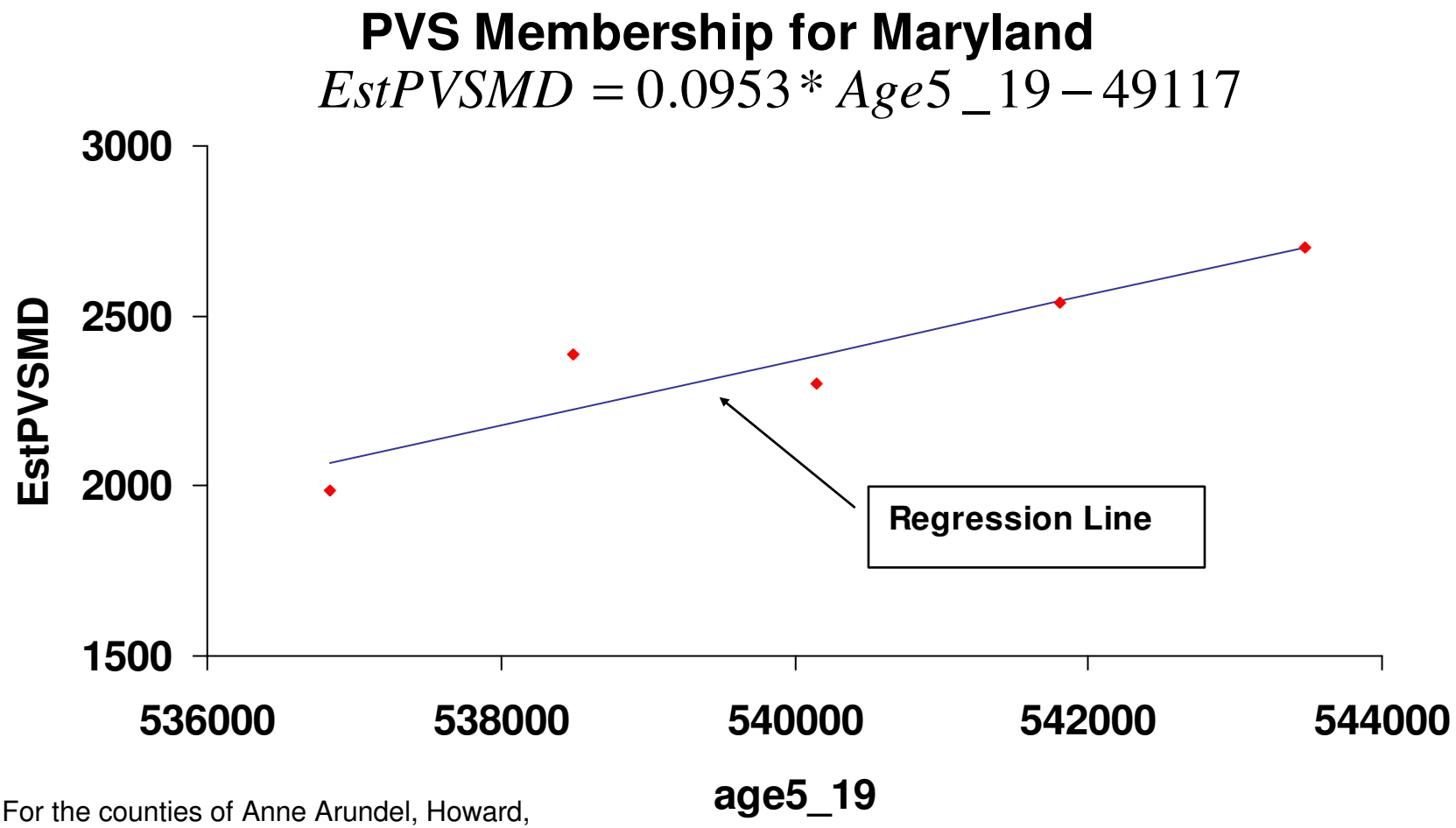
Aggregated population for the counties of Anne Arundel, Howard, Montgomery and Prince Georges.

Regression Model



For the cities of Alexandria, Fairfax, Falls Church and Manassas; and the counties of Arlington, Fairfax, Loudoun and Prince William.

Regression Model



For the counties of Anne Arundel, Howard, Montgomery and Prince Georges.

Regression Model

Model Performance								
		PATS Athletes	150					
		DC Athletes	280					
	Actual	Estimated	Actual	Estimated	Actual	Estimated	Relative	
Year	PVSVA	PVSVA	PVSMD	PVSMD	PVSALL	PVSALL	Error	
2002	3526	3568	1989	2067	5941	5916	0.42%	
2003	3935	3870	2385	2226	6739	6376	5.39%	
2004	4144	4173	2303	2384	6821	6836	0.23%	
2005	4508	4475	2539	2542	7452	7297	2.08%	
2006	4750	4777	2703	2700	7803	7757	0.59%	
2007	4940	5079	2896	2858	8217	8217	0.00%	
2008		5381		3017		8528		
2009		5683		3175		8988		
2010		5985		3333		9448		
(2002-2007)	EstPVSALL = EstPVSMD + EstPVSVA + DCAthletes							
(2008-2010)	EstPVSALL = EstPVSMD + EstPVSVA + DCAthletes - PATSAthletes							