# Development of a Spatially-Enabled Public-Use Database for End-Stage Renal Disease (ESRD) Policy Studies

Mark Stephens[1], Brendan-Maione-Downing[1], Sam Brotherton[1] and Matthew Gitlin[2]
[1]Prima Health Analytics, Boston MA, [2]InnoPeritus, Geneva CH

PRM36

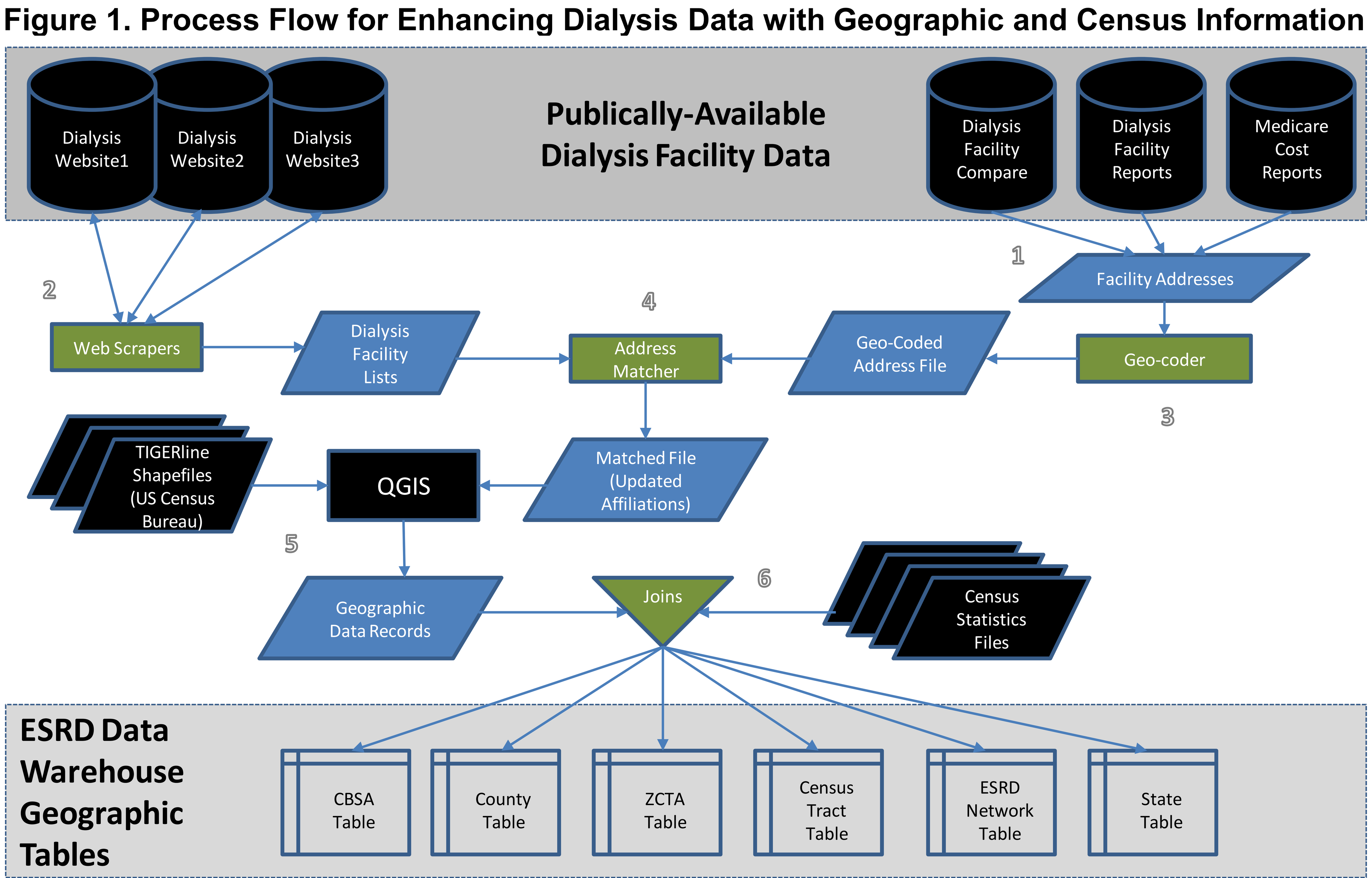
## BACKGROUND AND OBJECTIVES

While patients with ESRD represent only 1 percent of Medicare beneficiaries, this population accounts for 8 percent of total Medicare spending. The disproportionate cost of treating ESRD has resulted in a high level of monitoring and policy focus. A number of public databases currently exist for study of the ESRD program in the US, but these databases are difficult to integrate for use in policy analysis. The objective of this project was to develop a spatially-enabled, integrated ESRD "data warehouse" from free, publicly-available data sets, to enable a broadened scope of applications for policy analysis.

## METHODS

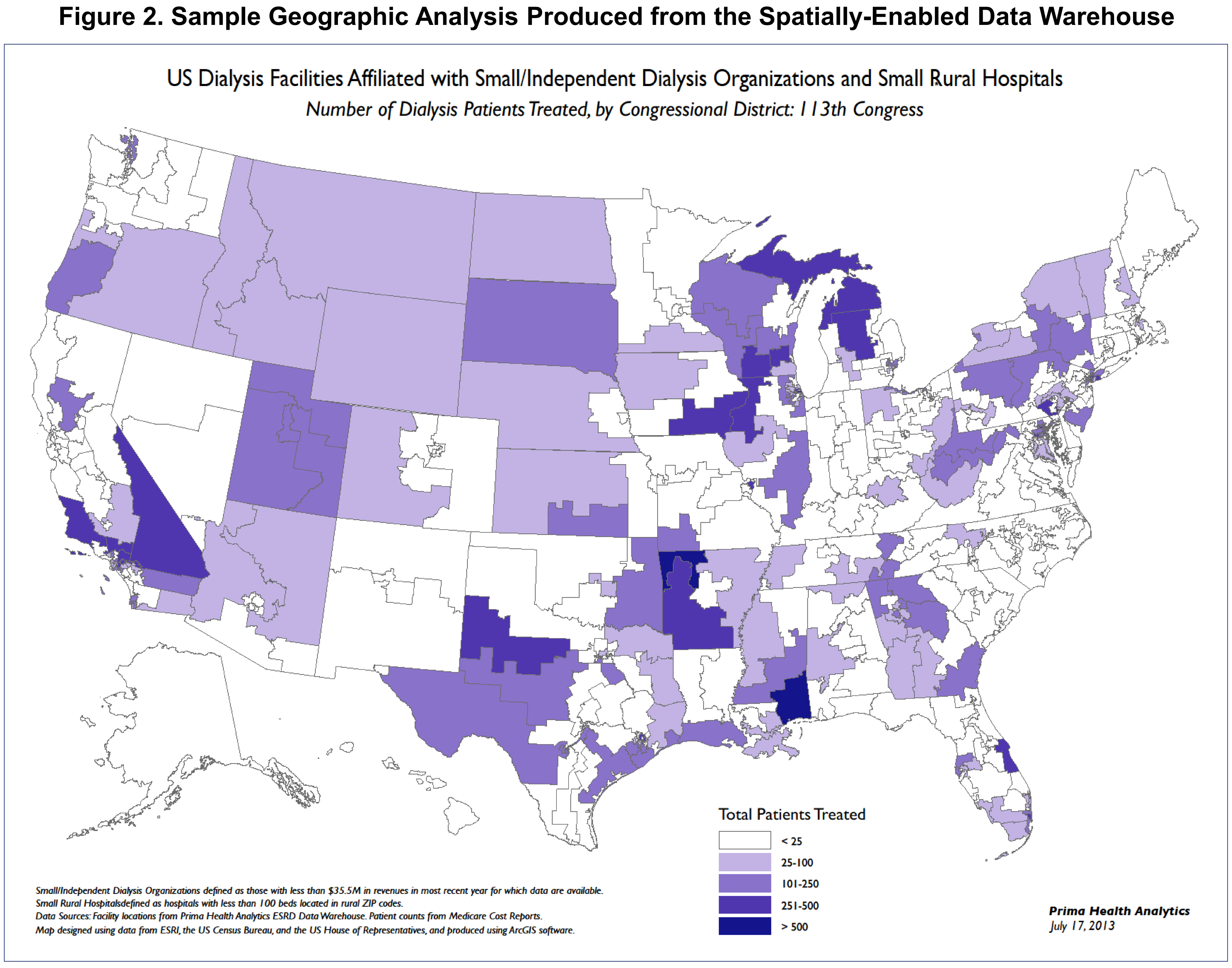
**Data Collection and Integration.** 1) ESRD-related datasets were collected from Medicare sources and integrated into a relational database management system (Figure 1). Datasets included the Dialysis Facility Compare database[1], dialysis facility and hospital Cost Reports by fiscal year[2], and Dialysis Facility Reports[3]. 2) Current dialysis facility listings were collected from websites of major dialysis chains using custom-built web-scraper software and/or manual lookup. 3) Using custom software, dialysis provider addresses were geo-coded to add latitude and longitude coordinates. 4) Using a custom address-matching algorithm, geo-coded address data were matched to the dialysis facility listings from websites to update facility chain affiliations. 5) Geo-coded and updated facility records were joined to spatial data files obtained from the U.S. Census Bureau website,[4] using QGIS software. This step added geo-political data including the Congressional District, ZIP-code Tabulation Area (ZCTA), census tract, and state and county FIPS codes of each facility's location. 6) This spatially-enabled output file was then joined to census statistics from the latest American Community Survey 5-year estimates[5] for each tract, ZCTA and county. Analysis sets, or "data marts" were then built to support specific research needs. Longitudinal datasets were created for trend analysis. Data are updated quarterly to keep the database current.

**Figure 1. Process Flow for Enhancing Dialysis Data with Geographic and Census Information**




**Challenges.** For a given entity, such as a dialysis facility, the raw source data files have disparate data structures requiring transformation into useable research datasets. Entity identifiers are inconsistent across sources, data file content and formats change from year to year, and there are missing data and inconsistencies in data values. Validation and outlier handling rules were therefore created for most data and were applied either during the build/update process or at the point of analysis. After transformations, there was a standardized data record for each facility, and the file format remains consistent over time.

**Figure 2. Sample Geographic Analysis Produced from the Spatially-Enabled Data Warehouse**




*Source: National Renal Administrators Association (NRAA) at http://www.nraa.org/files/Districts_Map_Based_on_Net_Revenue.pdf. Used by permission.*

## RESULTS

The resulting data warehouse contains over 6000 cost, utilization, quality, demographic and geographic variables, covering 100% of current Medicare ESRD providers (N=6568 facilities). Table 1 shows examples of the types of data included. Depending on data source, the database covers between 6 and 14 calendar years (2001-14). Quality of the data varies, but for most applications, outlier and missing observations are less than 10%.

**Table 1. Examples of Types of Data Included**

- ***ESRD Provider Costs, Revenues, and Service Volume***
  - Medicare Cost to Payment Ratios
  - FTEs by staff type
  - Drug costs and units
  - Dialysis Treatments by modality
- ***Patient Demographics And Characteristics***
  - Age, Race, Insurance Status
  - Dialysis Modality
  - Hemoglobin levels
  - Vascular Access Type
  - Average Number of Comorbidities
- ***Clinical and Process Outcomes***
  - Mortality, Hospitalization and Transplant Ratios
  - Dialysis Adequacy (URR, Kt/V)
  - Vaccinations
  - Pre-ESRD Treatment
- ***Facility Characteristics***
  - Firm/Chain Affiliation
  - Urban vs Rural location
  - Profit Status
  - Size (Stations, Patients)
  - Industry Segment (Hospital-based, Independent, etc)
- ***Population Demographics of Surrounding Community***
  - Racial composition
  - Income, Poverty Ratio
  - Minority composition
  - Educational Attainment
  - Age distribution

## DISCUSSION

Reporting and analysis using the integrated data sets is very easy using simple tools like Excel, or the data can be used in more complex applications such as GIS projects (Figure 2). Since implementation, project-specific data sets have been created to support several recently-published national policy studies on topics including patient access, provider efficiency, ESRD program costs and impacts of payment reform[6-9].

The national scope of the database can help decision-makers a) define the magnitude of a problem, b) identify where to focus efforts and c) set policy. It can provide an efficient and speedy method to acquire intelligence before investing in larger, more time consuming initiatives. Possible applications include clinical trial design, market access planning, product positioning and life cycle management.

## CONCLUSION

Public datasets of ESRD-related information can be integrated with population data and a GIS to support high-quality and cost-effective economic and clinical policy studies that would not otherwise be feasible.

## REFERENCES

1. Centers for Medicare & Medicaid Services. *Dialysis Facility Compare.* (2014).
2. Centers for Medicare & Medicaid Services. *Cost reports by fiscal year.* (2014).
3. University of Michigan Kidney Epidemiology and Cost Center. *Dialysis Facility Reports.* (2014).
4. US Census Bureau. *TIGER/Line Shapefiles Main Page - Geography*. (2014).
5. US Census Bureau. *American Factfinder Download Center*. (2014).
6. Shreay. *Health Serv. Res.* (2013).
7. Stephens. *J. Rural Health* 29, 339–348 (2013).
8. Brunelli. *Am. J. Kidney Dis.* 61, 947-956 (2013).
9. Stephens. *J. Health Econ. Outcomes Res.* (2013).

PRIMA Health Analytics

ISPOR 19TH ANNUAL INTERNATIONAL MEETING, MONTREAL, QC CANADA, MAY 31 – JUNE 4, 2014