arcos and arcospy: R and Python packages for accessing the DEA ARCOS database from 2006 - 2014

Steven Rich[1], **Andrew Ba Tran**[1], **Aaron Williams**[1], **Jason Holt**[1], **Jeffery Sauer**[2], **and Taylor M. Oshan**[2]

1 The Washington Post **2** University of Maryland, College Park, Department of Geographical Sciences

Summary

In the early 2000s governmental agencies across the United States began to observe increases in the number of all-cause opioid-involved deaths (Stopka et al., 2019a). The Centers for Disease Control (CDC) describe this Opioid Overdose Epidemic as occuring in three waves, with the first wave attributed to the widespread distribution of prescription opioids, reaching an opioid prescribing rate as high as 81.3 per 100 persons in 2012 ("Prescribing practices," 2019). While the more recent wave two and wave three of the Opioid Overdose Epidemic are largely attributed to illicit opioids like heroin and fentanyl, prescribing rates and prescription misuse remain high (Guy et al., 2017; Ko et al., 2020).

Researchers, journalists, and government agencies are still actively investigating the myriad impacts of prescription opioids on the Opioid Overdose Epidemic. One powerful tool for understanding prescription opioid distribution is the Drug Enforcement Administration's (DEA) Automation of Reports and Consolidated Orders System (ARCOS). ARCOS tracks the commercial distribution of controlled substances in the United States, including opioid analgesics. ARCOS data is highly detailed, tracking commercial origin, pharmacy order frequency, point-of-sale distribution, and more. For a variety of reasons ranging from patient confidentiality to protecting trade secrets, access to sub-state ARCOS data is available only for approved requests (e.g. research or litigation) (Grubbs, 2014). Recent litigation efforts by *The Washington Post*, HD Media, and local journalists allowed for the public release of an anonymized, large portion of the ARCOS database from 2006 to 2012, with additional data for 2013 and 2014 now also available. `arcos` and `arcospy` are open-source API wrappers in `R` and `Python`, respectively, that allow researchers and interested citizens to easily access this newly available portion of the ARCOS database.

Statement of Need

Previously, researchers wanting to use ARCOS data relied on what was made available by the DEA, typically in the form of state-level estimates, or submitted special access requests to the DEA (Kenan, Mack, & Paulozzi, 2012; Reisman, Shenoy, Atherly, & Flowers, 2009). While alternative data on prescription records are offered by the Centers for Medicare & Medicaid Services in the Medicare Provider Utilization and Payment Datasets, this data pertains to a specific sample of the population and spans a different set of years (2011 to 2017). In addition, the level of detail in the ARCOS data offers substantial opportunity for commodity analysis about prescription opioids including market dynamics, product demand, supply chain flow, and more. This relationship between commercial distribution and the Opioid Overdose Epidemic is an important area for future study to define and recognize the warning signs of potentially problematic prescribing practices (Van Zee, 2009). The release of national,

longitudinal, sub-state ARCOS data is a major contribution for researchers interested in the distribution of prescription opioids and the subsequent sociomedical impacts.

In raw format, the ARCOS database is more than 130 gigabytes and includes several hundred columns. Thus, the purpose of `arcos` and `arcospy` are meant to:

- Simplify access to an open, large, robust prescription opioid database
- Provide measures of prescription opioid distribution relevant to both the medical and social sciences
- Promote analytical flexibility and reproducibility through mirrored functionality across `R` and `Python`

API Structure

The `arcos` and `arcospy` API is publically available and hosted using the OpenAPI specification. The primary maintainers of API database are members of the Data Reporting Team at *The Washington Post*. A key is required to use the API. The standard key is `WaPo` and additional keys may be sourced from the Github repository. Guidelines on using the API are available from *The Washington Post*.

All commands share the same name between `arcos` and `arcospy`. This allows users to easily switch between languages if the need arises. Outputs from all of the functions are delivered in popular formats - `data.frames` in R and `pandas.DataFrame` in python - to enable statistical, spatial, network, or other types of analysis.

Both `arcos` and `arcospy` use parameter delivery - `urltools` in R and `requests` in Python - to build the API query. Checks are in place to ensure that invalid inputs are not passed to the API. For example, a series of integers cannot be passed as a county name. Corrective warning messages are returned to users who provide invalid inputs.

Data Availability and Basic Usage

Data can be gathered at the pharmacy, distributor, county, or state as the geographic unit of analysis. Depending on the geographic level, there may be raw, summarized, or supplemental data available. For example, the `county_raw()` command returns each individual ARCOS record for a given county from 2006 to 2014. The following code chunk demonstrates this function in `R`:

```r
library(arcos)
# Gather all ARCOS records for Hill County, Montana
HillRaw <- county_raw(county = "Hill",
                      state = "MT",
                      key = "WaPo")
head(HillRaw)
```

REPORTER_DEA_NO	REPORTER_BUS_ACT	REPORTER_NAME	...	dos_str
PM0023046	DISTRIBUTOR	MCKESSON CORPORATION	...	5.0
PM0023046	DISTRIBUTOR	MCKESSON CORPORATION	...	5.0
PM0023046	DISTRIBUTOR	MCKESSON CORPORATION	...	5.0
PM0023046	DISTRIBUTOR	MCKESSON CORPORATION	...	5.0
PM0023046	DISTRIBUTOR	MCKESSON CORPORATION	...	7.5
PM0023046	DISTRIBUTOR	MCKESSON CORPORATION	...	20.0

However, the `summarized_county_annual()` command returns the annual summarized totals for a given county for each year of 2006 to 2012. The following code chunk demonstrates this function in `Python`:

```python
from arcos import summarized_county_annual
# Gather summarized ARCOS records for Hill County, Montana
HillSummarized = summarized_county_annual(county = "Hill",
                                          state = "MT",
                                          key = "WaPo")
HillSummarized.head()
```

	BUYER_COUNTY	BUYER_STATE	year	count	DOSAGE_UNIT	countyfips
0	HILL	MT	2006	1516	594700	30041
1	HILL	MT	2007	1710	505430	30041
2	HILL	MT	2008	2467	715560	30041
3	HILL	MT	2009	3200	851560	30041
4	HILL	MT	2010	3290	803760	30041

Given the number of records, users should anticipate that commands querying for raw data will take longer than commands querying for summarized data. Full documentation on how data is collected by the DEA is available in the ARCOS Registrant Handbook. `arcos` and `arcospy` also include supplemental commands that return relevant auxiliary data - such as county population - gathered from the American Community Survey. A description of each of the functions currently offered, as well as examples in `R` and `Python` demonstrating functionality, are available on the shared `arcos` and `arcospy` repository.

There are several ways to conceptualize the unit of analysis for opioids from the present data. These include the total number of records, the total number of all opioid pills, the total number of specific opioid pills (i.e. oxycodone versus hydrocodone), or the total amount (in weight) of all or specific opioid pills. Other common units of analysis that may be of interest include morphine milligram equivalents (MMEs) or prescription counts (Stopka et al., 2019b). Users should choose a unit of analysis that has precedent in their discipline and take appropriate steps to standardize the data (e.g. by population or another stratum) when necessary.

Conclusion

`arcos` and `arcospy` allows access to a substantial amount of previously unavailable data on prescription opioid distribution in the United States during the years leading up to the present Opioid Crisis. Data from the DEA ARCOS system has been used in scientific publications, primarily at the intersection of health and criminology, to investigate trends in analgesic use and potential abuse (Gilson, Ryan, Joranson, & Dahl, 2004; Joranson, Ryan, Gilson, & Dahl, 2000). Additionally, the data made available by `arcos` has been used extensively by journalists at *The Washington Post* and local news outlets to report on trends in prescription opioid distribution (Diez, 2019; Top, 2020). ARCOS data can be merged (non-spatially or spatially) with other United States statistical products through packages like `tidycensus` in `R` and `cenpy` in `Python`, opening numerous doors for research and teaching exercises. Examples of these possibilities are available on the Github repositories for `arcos` and `arcospy`. The flexibility to query the ARCOS DEA database using commands of the same name enhances reproducibility across languages and ease of access. Expanding the ways in which researchers and journalists can analyze robust datasets like ARCOS is an important step towards understanding how the United States arrived at the present Opioid Overdose Epidemic.

Availability

`arcos` is available on CRAN as well as Github.

`arcospy` is available on PyPI as a pip installable package as well as Github.

The repository for this article and additional information is stored on the shared the shared `arcos` and `arcospy` repository on Github.

Citations

Diez, M. (2019). The post released the dea's data on pain pills. Here's what local journalists are using it for. *The Washington Post*. Retrieved from https://www.washingtonpost.com/national/2019/08/12/post-released-deas-data-pain-pills-heres-what-local-journalists-are-using-it/&usg=AOvVaw3TrlkTVOxr_jFBdiKIU61G

Gilson, A., Ryan, K., Joranson, D., & Dahl, J. (2004). A reassessment of trends in the medical use and abuse of opioid analgesics and implications for diversion control: 1997-2002. *Journal of Pain and Symptom Management*. doi:10.1016/j.jpainsymman.2004.01.003

Grubbs, P. (2014). Privacy impact assessment for the registrant information consolidated system. *Drug Enforcement Administration*. Retrieved from https://www.dea.gov/sites/default/files/2018-06/rics_%20pia_060414.pdf

Guy, G. P., Zhang, K., Bohm, M. K., Losby, J., Lewis, B., Young, R., Murphy, L. B., et al. (2017). Vital signs: Changes in opioid prescribing in the united states, 2006–2015. *Morbidity and Mortality Weekly Report*. doi:10.15585/mmwr.mm6626a4

Joranson, D., Ryan, K., Gilson, A., & Dahl, J. (2000). Trends in medical use and abuse of opioid analgesics. *Journal of the American Medical Association*. doi:10.1001/jama.283.13.1710

Kenan, K., Mack, K., & Paulozzi, L. (2012). Trends in prescriptions for oxycodone and other commonly used opioids in the united states, 2000-2010. *Open Medicine*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3659213/

Ko, J. Y., D'Angelo, D. V., Haight, S. C., Morrow, B., Cox, S., Salvesen von Essen, B., Strahan, A. E., et al. (2020). Vital signs: Prescription opioid pain reliever use during pregnancy — 34 u.s. Jurisdictions, 2019. *Morbidity and Mortality Weekly Report*. doi:10.15585/mmwr.mm6928a1

Prescribing practices. (2019).*Centers for Disease Control and Prevention, National Center for Injury Prevention and Control*. Retrieved from https://www.cdc.gov/drugoverdose/data/prescribing/prescribing-practices.html

Reisman, M., Shenoy, P., Atherly, A., & Flowers, C. (2009). Prescription opioid usage and abuse relationships: An evaluation of state prescription drug monitoring program efficacy. *Substance Abuse*. doi:10.4137/sart.s2345

Stopka, T., Amaravadi, H., Kaplan, A., Hoh, R., Bernson, D., Chui, K., Land, T., et al. (2019a). Opioid overdose deaths and potentially inappropriate opioid prescribing practices (pip): A spatial epidemiological study. *International Journal of Drug Policy*. doi:10.1016/j.drugpo.2019.03.024

Stopka, T., Amaravadi, H., Kaplan, A., Hoh, R., Bernson, D., Chui, K., Land, T., et al. (2019b). Opioid overdose deaths and potentially inappropriate opioid prescribing practices (pip): A spatial epidemiological study. *International Journal of Drug Policy*. doi:10.1016/j.drugpo.2019.03.024

Top, E. (2020). Suburbs saw large influx of prescription opioids. *Bowie Blade-News.* Retrieved from https://baltimorefishbowl.com/stories/analysis-marylands-suburbs-saw-a-large-influx-of-presc

Van Zee, A. (2009). The promotion and marketing of oxycontin: Commercial triumph, public health tragedy. *American journal of public health.* doi:10.2105/AJPH.2007.131714