Measuring Product Innovation in the Pharma Industry Using Nontraditional Data Sources

Eirik Iversen (VT), Luke Kim (VT), Raghav Sawhney (VT), Fangfang Lee (NYU) and Bayoán Santiago Calderón (CGU)

SDAL: Gizem Korkmaz, Ian Crandell, and Stephanie Shipp

Sponsor: Gary Anderson, The National Center for Science & Engineering Statistics (NCSES) at the National Science Foundation (NSF)

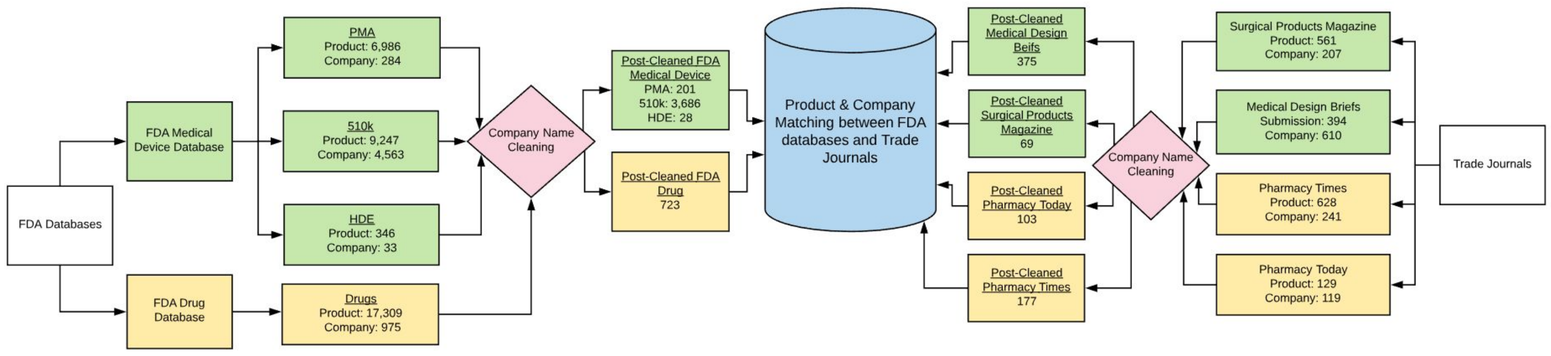
Introduction

- Innovation is traditionally measured through surveys of selected companies (e.g., Business R&D and Innovation Survey (BRDIS)). While BRDIS measures innovation incidence, i.e., the number of innovating firms, NCSES is interested in exploring the possibility of using *non-traditional data* to enrich and complement innovation measures. This project aims to test the feasibility of developing methods to measure business innovation using *non-traditional data sources* focusing on *opportunity and administrative data* (e.g., product announcements, press releases, websites and other information obtained through web scraping or queries of selected companies).
- Our current focus is on Pharmaceutical Drugs and Medical Devices that are heavily regulated by the Food and Drug Administration (FDA). This is appealing because there is a lot of public information about these products on FDA sources. We identify the industries by their 4-digit NAICS (North American Industry Classification System) codes:
  - ❏ *3254*-Pharmaceutical and Medicine Manufacturing & *3391*-Medical Equipment and Supplies Manufacturing
- The goal is to match the product approvals documented in FDA databases to new product announcements in all data sources to compare to the new products covered in trade journals.

Data and Methods

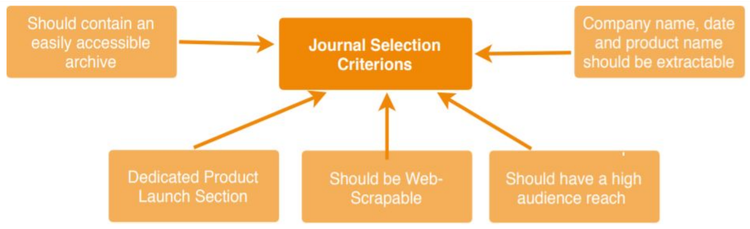
FDA Databases

- ❏ Drugs @ FDA website provides a comprehensive list of all drug approvals up-to-date.
- ❏ Medical Devices are classified by the level of risk on the user/patient by the FDA:
  - PMA (Pre-Market Approval): Includes Class III devices which have the highest risks and need the most examination. It takes a long time to process a PMA and is usually the most expensive approval process.
  - 510k (Pre-Market Approvals): Includes Class II devices that have moderate risk and are approved by being "Substantially Equivalent" to devices that have previously been approved.
  - HDE (Humanitarian Device Exemption): This is the process where medical devices are allowed to be marketed without requiring evidence of effectiveness.
    *Note: Class 1 devices do not need FDA approval in most cases.



Trade Journals

- As a part of our trade journal discovery, we studied similar research, e.g., [1] and [2], to obtain a list of trade journals, to understand their methods and criteria for selecting the relevant journals.
- We selected 4 trade journals from a larger compiled list to be web-scraped for product innovations based on the criterions given on the right.
  - Drugs: Pharmacy Times and Pharmacy Today
  - Medical Devices: Medical Design Briefs and Surgical Products Magazine

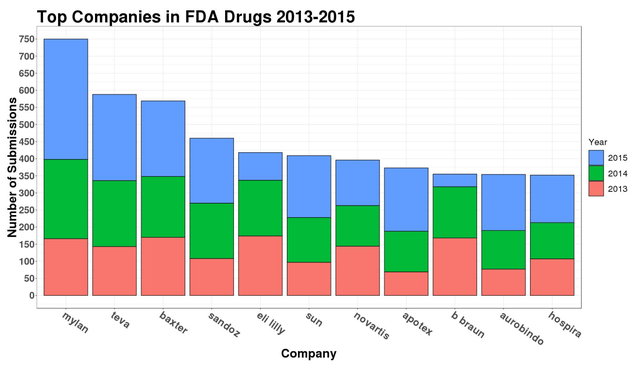


- We used the Levenshtein Distance method [3] to identify name standardization issues and created a set of rules and applied them to all data sources.
- Challenges
  - Company names can vary across data source
    Further complicated by parent-subsidiary relationships, mergers, company divisions and cooperative projects.

Findings

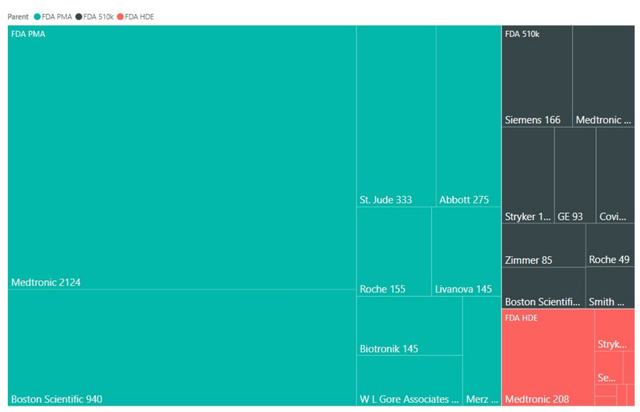
FDA Drug Database (2013 - 2015)

- 17,309 drug approvals (1,820 original submissions) and 723 companies.
- Top 10 companies with the highest number of approvals are given below.



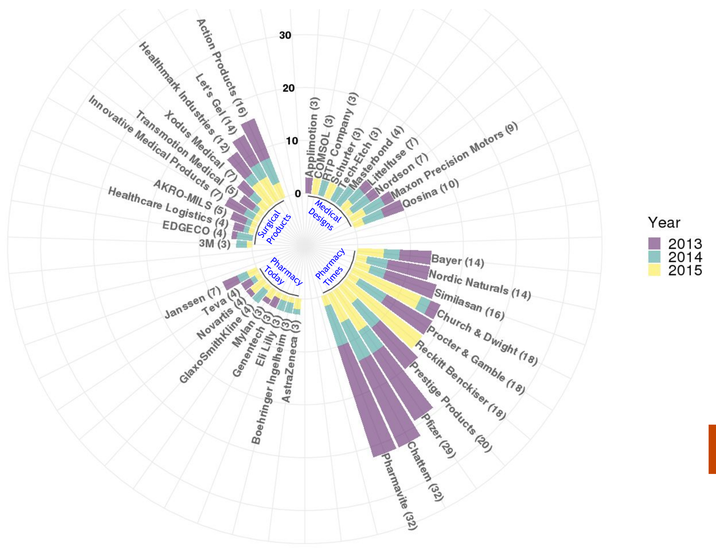
FDA Medical Device Databases

- Top companies for each submission type (indicated by different colors). The approval counts are given next to the company names.

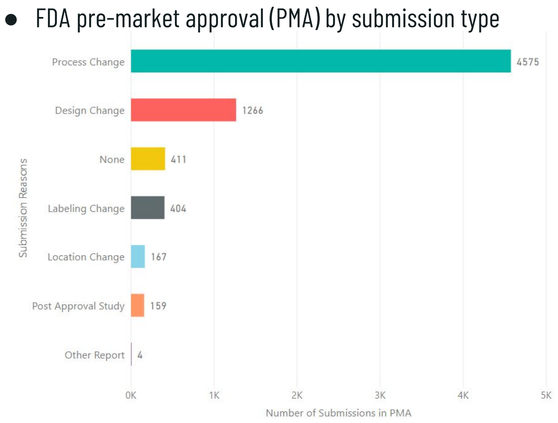


Trade Journals

Top published companies by trade journal



- FDA pre-market approval (PMA) by submission type



Matching Companies in Data Sources

Number of unique companies in each data source (2013-2015)

Data Source	Number of companies (after cleaning)	Number of companies matching FDA	Total number of FDA approvals for these companies	Total number of new products in trade journal for these companies (% FDA approval)
FDA Medical Device Database	3,915			
Medical Design Briefs	375	13	384	17 (4.43%)
Surgical Products	69	17	344	29 (8.43%)
FDA drugs	723			
Pharmacy Times	177	68	4,950	249 (5.03%)
Pharmacy Today	103	57	6,328	91 (1.44%)

We have identified the above company matches across these data sources. The goal is to identify new product announcements that are covered in trade journals in the FDA drug and medical device approval databases.

Next Steps

- Obtain product approval counts by submission and supplement types as some of these types (e.g., location change) will not count as innovation, hence may not be announced in the trade journals.
- Match *at the product level* and and aggregate by company and industry
- Develop more advanced methods for company name standardization and acquire new data sources for more specific information on industry and companies (e.g., subsidiaries, revenue, size)
- Identify additional trade journals from the literature and based on the industry experts' uses
- Introduce additional data sources: SEC filings, press releases from company websites
- Extend the methods to other industries (e.g., automotive, technology)

References

[1] Edwards, K. L., & Gordon, T. J. (1984). Characterization of innovations introduced on the US market in 1982. *Futures Group; Reproduced by NTIS.*

[2] Acs, Z. J., & Audretsch, D. B. (1988). Innovation in large and small firms: An empirical analysis. *The American Economic Review, 678-690.*

[3] Levenshtein, Vladimir I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*. 10 (8): 707–710.

bi.vt.edu/sdal

Stay Connected: *facebook.com/biocomplexity*
twitter.com/biocomplexityvt
youtube.com/c/biocomplexity
google.com/+biocomplexity