Approaches to Defining Topic-Related Portfolios of Biomedical Research

28th Annual Conference of the American Evaluation Association
October 15-18, 2014

Session RTDE1
October 16

NCI Office of Science Planning and Assessment
James Corrigan, Ph.D.
Danielle Daee, Ph.D.
Elizabeth Hsu, Ph.D., M.P.H.
Examples of Issues Addressed by Defining Portfolios

- A common task within the context of NIH evaluations / assessments for a variety of purposes, such as:
  - Describing the investment in a particular topic area

- Drilling down to describe characteristics of projects in a topic area in terms of:
  - Funding mechanisms;
  - Funding sources;
  - Principal Investigator characteristics;
  - Research institution characteristics

- Determining portfolio gap areas

- Providing an evidence base for initiative launch / renewal decisions

- Defining comparison groups of projects focused on similar topics
Three Approaches to Defining Portfolios

1. TEXT MINING
   PROSPECTIVE
   automated comparison of proposal text to expert-defined categories

2. EXPERT REVIEW
   PROSPECTIVE
   entire research proposal read

3. PUBLISHED RESULTS
   RETROSPECTIVE
   topic relevance defined by keyword search of published results
Approach 1. Text Mining

NCI Pediatric portfolio defined by:

- **RCDC categories:** Pediatric, Pediatric AIDS, Pediatric Research Initiative (combined with ‘OR’)
  
  RCDC: Research, Condition, and Disease Categorization
  
  - NIH-level coding of projects in reportable categories
  
  - text mining comparison of proposal content to sets of weighted expert-identified concepts & terms that define each category.

- plus keyword “Cancer” (free text search of Abstract, Aims, Title, Summary Statement)

- limited to FY 2011 NCI awarded projects

- note: Data for intramural and contract awards is not comprehensive in IMPAC II

Data Source: IMPAC II

Search Interface: Query, View, Report (QVR)
NCI Pediatric Portfolio defined by:
- Special Interest Category: **Childhood Cancers**
  - Special Interest Categories (SICs)
    - NCI coding staff analyze grant applications, research contracts, and intramural projects to classify each project to different SICs
    - Coding staff assign % relevance (range 0-100%) for some non-mutually exclusive SICs
  - limited to FY 2011 NCI awarded projects
- note: comprehensive intramural and contract award information is available in NFRP
NCI FY11 Pediatric Cancer Portfolio
Prospective Approach – Research Proposal

1. TEXT MINING PROSPECTIVE
RCDC Pediatric + cancer NCI FY 2011 Projects
n = 196

2. EXPERT REVIEW PROSPECTIVE
SIC Childhood Cancer NCI FY 2011 Projects
n = 674

Intersection
n = 133

63 unique records identified by the RCDC approach

541 unique records identified by expert review

Total portfolio: 737 FY11 projects
• Caveats of text mining and expert review approaches:
  – based on portions of proposal text
  – prospective coding will not capture
    • areas not explicitly described in the proposal
    • studies that shift focus as research progresses
      – aims may change
      – findings may have unanticipated implications for areas
        not addressed in original research proposal
Three Approaches to Defining Portfolios

1. TEXT MINING
   PROSPECTIVE
   automated comparison of proposal text to expert-defined categories

2. EXPERT REVIEW
   PROSPECTIVE
   entire research proposal read

3. PUBLISHED RESULTS
   RETROSPECTIVE
   topic relevance defined by keyword search of published results
Defining Portfolios by Published Research
Retrospective Approach

Research Proposal → Research → Research Outputs

Are “proposed results” and “actual results” relevant to the same topics?

not necessarily if the projects:
- have changed scope
- have a general, basic, or adult cancer focus that have made unanticipated discoveries relevant to pediatric cancer
- did not exhaustively describe relevance in the project proposal
Defining Portfolios by Published Research
Retrospective Approach

3. PUBLISHED RESULTS
RETROSPECTIVE

- topic relevance defined by keyword search of published results

Research Outputs

- Identify pediatric cancer-relevant published results by keyword search

Identify the NCI awards acknowledged by these publications

What fraction of the awards that contribute to discoveries in pediatric cancer research are actually captured by the prospective approaches?

General Approach:
“Grant Number” field from PubMed pediatric cancer search results linked to FY11 NCI awarded projects

- PubMed Search: "NCI" OR "CA"[Grant Number]) AND (pediatric OR childhood) AND (cancer), filter for 2012

- Project (grant) numbers are linked to other defined NCI portfolios for FY11 for analysis
  - text mining-defined Pediatric Portfolio
  - expert review-defined Pediatric Portfolio
  - FY11 NCI Portfolio, all funded projects

- Caveats:
  - Conservative Estimate:
    - intramural or contract awards not captured
    - publication lag (new projects in FY11 likely haven’t published by 2012)
  - dependent on accurate research support data
NCI FY11 Pediatric Cancer Portfolio
Multiple Approaches

1. TEXT MINING
   PROSPECTIVE
   RCDC Pediatric + cancer
   n = 196
   47 unique to this approach

2. EXPERT REVIEW
   PROSPECTIVE
   SIC Childhood Cancer
   n = 674
   471 unique to this approach

3. PUBLISHED RESULTS
   RETROSPECTIVE
   n = 368
   175 unique to this approach (47%)

Total portfolio: 912 FY11 projects
47% of projects identified by published research approach are not coded to Pediatric.
Defining Portfolios by Published Research

Pediatric Portfolio

Major Disease Sites
FY11 NCI Projects Linked to Pediatric Publications

Projects NOT CODED to Pediatric
n = 175

- General Cancer: 34%
- Lung: 9%
- Kidney: 1%
- Gastrointestinal: 7%
- Multiple Major: 5%
- Sarcoma: 3%
- Brain/CNS: 1.0%
- Leukemia/Lymphoma: 11%
- Other: 11%

Projects CODED to Pediatric
n = 193

- Leukemia/Lymphoma: 31%
- Breast: 12%
- General Cancer: 12%
- Brain/CNS: 19%
- Other: 3%
Defining Portfolios by Published Research
Pediatric Portfolio

Activity Codes
FY11 NCI Projects Linked to Pediatric Publications

Projects NOT CODED to Pediatric
n = 175

Projects CODED to Pediatric
n = 193
Defining Portfolios by Published Research

Retrospective Approach

Summary of Retrospective Approach based on Published Research

- Approach allowed expansion of portfolio to capture published results relevant to pediatric cancers, including projects with:
  - basic or adult cancer research focus with unanticipated implications for pediatric cancers
  - generalized project proposals like center grants and training grants (P, T)

- Caveats
  - Could suggest some over-reporting of grant funding in publication acknowledgments

- Next Steps
  - A closer look at why acknowledged projects were not coded to pediatric:
    - a more detailed manual review or text similarity approach comparing publications and research proposals
  - Looking at alternate prospective & retrospective strategies
Acknowledgements

• NCI Office of Science Planning and Assessment
  – Danielle Daee
  – Betsy Hsu
  – Larry Solomon
  – Karl Poonai
  – Vivian Horovitch-Kelley

• NCI Research Analysis and Evaluation Branch
  – NCI Funded Research Portfolio
1. TEXT MINING
   PROSPECTIVE
   RCDC Pancreatic Cancer
   n = 302
   88 unique to this approach

2. EXPERT REVIEW
   PROSPECTIVE
   SITE Pancreatic
   n = 528
   252 unique to this approach

PUBLISHED RESULTS
   RETROSPECTIVE
   keyword pancreatic AND cancer
   n = 770
   538 unique to this approach (70%)

Total portfolio: FY11 projects
Defining Portfolios by Published Research
Pancreatic Portfolio

FY11 NCI Projects Linked to Pancreatic Cancer Publications
n = 770

coded to Pancreatic 30%
not coded to Pancreatic 70%