Data Extraction and Quality Challenges/ Lessons Learned

Keith Marsolo, PhD
Duke University School of Medicine

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Panelists

- Dana Dailey, PhD
  - Fibromyalgia TENS in Physical Therapy Study (FM-TIPS)
- Shruti Gohil, MD
  - INtelligent Stewardship Prompts to Improve Real-time Empiric Antibiotic Selection for Patients (INSPIRE)
- Corita Grudzen, MD
  - Primary Palliative Care for Emergency Medicine (PRIM-ER)

Session Goals

- Learn about issues encountered with EHR integration, data extraction, and data quality
- Share ways to mitigate data-related challenges, such as implementation monitoring and strong cross-collaborative team structures

FM-TIPS: Lessons Learned

Data Extraction and Quality

An Embedded Pragmatic Clinical Trial in Physical Therapy (PT) Clinics

Dana Dailey, PT, PhD

FM-TIPS
Fibromyalgia TENS In Physical Therapy Study

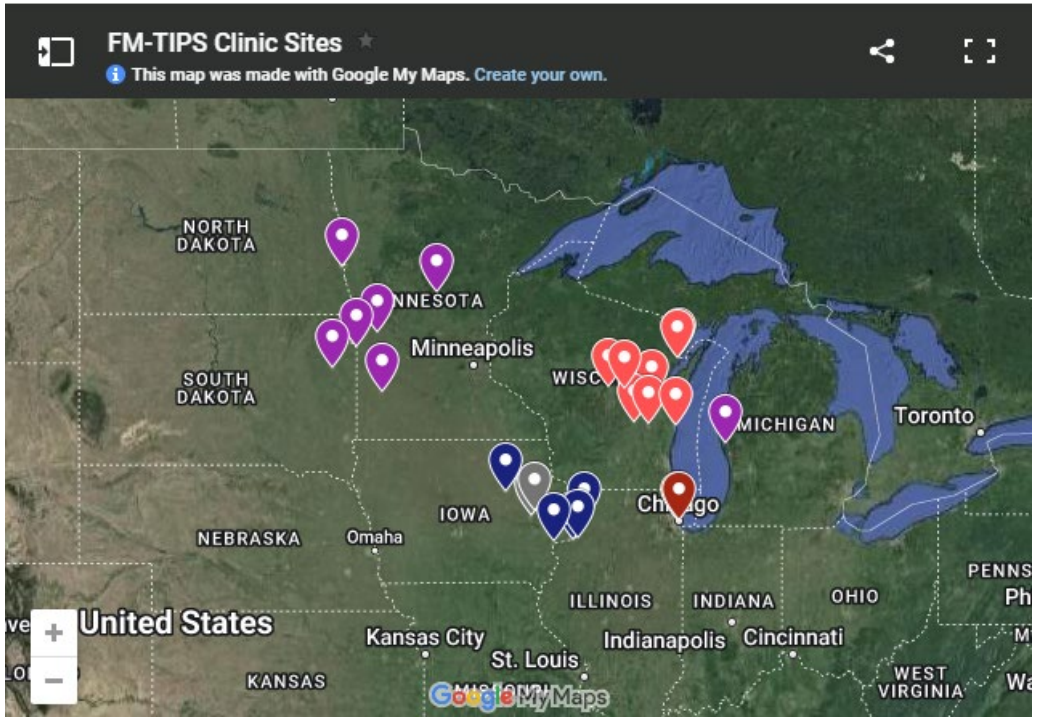
Acknowledgements: Research supported by National Institutes of Health Grant UG3/UH3 AR076387-01 and UL1TR002537; NIH Collaboratory

FM-TIPS
Fibromyalgia TENS In Physical Therapy Study

FM-TIPS Study Overview

The goal of FM-TIPS is to test the feasibility and effectiveness of adding transcutaneous electrical nerve stimulation (TENS) to standard physical therapy (PT) care in a real-world physical therapy setting

- 25 active physical therapy clinics
- 100+ physical therapists
- 5 active healthcare systems
- 11 EHRs



FM-TIPS Data Overview

FM-TIPS
Fibromyalgia TENS In Physical Therapy Study

REDCap: Participant Entered, Primary & Secondary Outcomes

EHR: Therapist Entered, Secondary Outcomes

TENS Usage Data

Episode of Care

Study Flag, Name, DOB, Site, Start/End Dates, ICD 10 codes, # visits attended, # visits scheduled

Visit Level

Study Flag, Name, DOB, Site, Visit Date, CPT Codes, *PSFS

*PSFS: Patient reported outcome measure initially done with therapist and moved to REDCap

1 Lesson 1: Initial planning – do your homework

FM-TIPS
Fibromyalgia TENS In Physical Therapy Study

At the grant writing stage

- Consult with data managers about your data and the process
- Consult with multiple clinics/providers about the EHR and data extraction process
- Include data extraction costs in your budget
  - Setting up the process on the clinic side cost
  - Periodic extraction cost
  - Cleaning, mapping, and reporting of data

2 Lesson 2: Clinicians need support

FM-TIPS
Fibromyalgia TENS In Physical Therapy Study

- Limited Time
  - PT's who manage data extraction are focused on patient care
  - Building relationships with clinicians can be challenging
    - Limited availability of the clinician
    - Multiple people involved in the process of data extraction (clinicians, coders, billers, etc.)
    - Starting a study during a pandemic is hard
- Limited Experience
  - The people who manage data extraction have limited experience in data extraction for research
  - Limited experience in data transfer
- Limited Funds
  - Each healthcare system has different costs associated with data extraction
  - The study may have limitations in funding
  - Implications for patient care

3 Lesson 3: Data managers are integral to the process

FM-TIPS
Fibromyalgia TENS In Physical Therapy Study

- Data collection
- Data extraction
  - Timeframes
  - Data
- Integration of extracted data
  - Matching data
  - Coding data
  - Interpreting data
  - Reporting of the data

4 Lesson 4: It always takes longer than you think

FM-TIPS
Fibromyalgia TENS In Physical Therapy Study

- Developing relationships
- Learning curve for the clinics and clinicians
- Transfer of information – method, process, firewalls, etc.
- Matching and coding of data
- Invoicing for data extraction

2021	• Healthcare System 1-5 Activated • First Participant February
2022	• Data transfers began • Healthcare System 6 Activated
2023	• Healthcare System 4 Deactivated • Regular data transfers
2024	• Regular data transfers
2025	• Last data transfer

FM-TIPS
Fibromyalgia TENS In Physical Therapy Study

Summary

- Data extraction is a challenging process
- No two EHR's are the same
- Clinicians who manage data extraction are busy
- Data managers and clinicians are crucial to the process of data extraction
- Develop relationships with the healthcare systems and clinicians
- Be patient and be persistent

FM-TIPS Team

INSPIRE Abdominal & Skin/Soft Tissue Infection Trials

INtelligent Stewardship Prompts to Improve Real-time Empiric Antibiotic Selection for Patients

NIH Collaboratory In-Person Steering Committee Meeting
Data Extraction and Quality Challenges Panel
May 10, 2024

Shruti K. Gohil, MD, MPH
Assistant Professor, Division of Infectious Diseases
Associate Medical Director, Epidemiology & Infection Prevention
University of California, Irvine School of Medicine

INSPIRE Trials: Purpose & Design

- **Purpose:** Reduce unnecessary empiric broad-spectrum antibiotic use
- **Design:** Cluster-randomized trials, 92 HCA Healthcare hospitals, non-ICU patients
- **Intervention:** CPOE prompts for abdominal or skin/soft tissue infections
- **Outcomes:**
  - **Effectiveness** – antibiotic use first 3 inpatient days
    - Primary – any broad-spectrum antibiotics
    - Secondary – antibiotic subsets
  - **Safety:** days to ICU transfer, hospital length of stay

Multidrug-Resistant Organism (MDRO) Models

Challenges in curating data during pre-trial modeling

- Extensive data streams and cleaning
- >60 variables from 140 hospitals > 500,000 patients across 3 years
- 10 separate models to predict MDRO infection risk

Lessons learned

- Pre-trial data helpful for revealing complexities
- Needs realistic budgeting of analytic time
- Pivoted to pull and clean data throughout trial

Speeding Up Data Cleaning

Challenges in data alignment - despite a central data repository

- Many to one matching for medication or lab data
- Data standards exist but user overrides occurs
- Variation in hospital or provider order sets

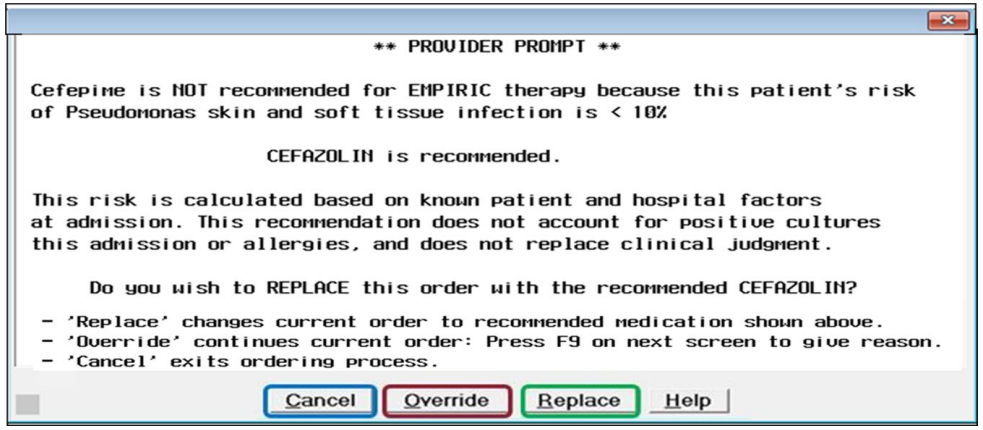
Lessons learned: proactive steps

- Smart sampling for layering data checks
- Monthly hospital data reports
- Monthly investigations and cleaning

Data From CPOE Prompts

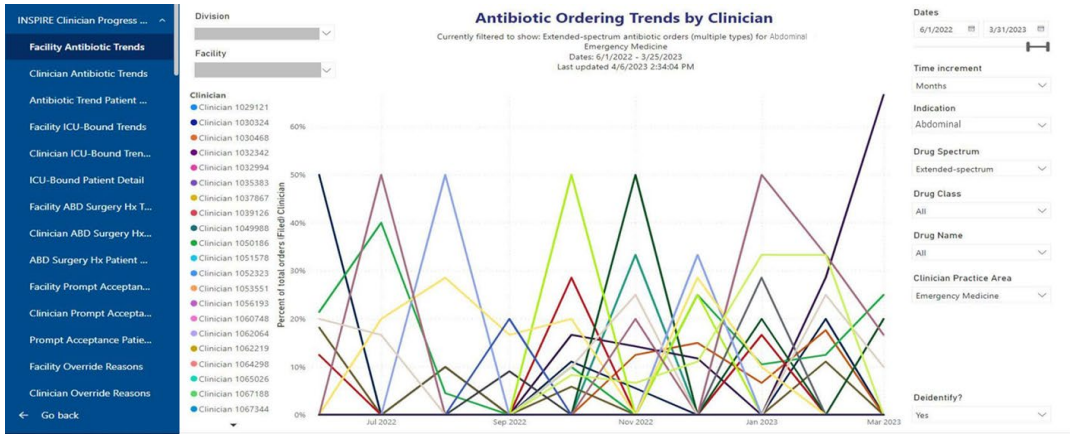
Build prompt *de novo*

- Specifications and queuing take time
- Capture prompt responses
- Feedback reports



Lessons learned

- Anticipate queuing time
- Leverage health system partners for support
- Allow health system solutions



Speeding Up Trial Analysis

Challenges in analysis

- Multiple models – effectiveness vs safety
- Hierarchical clustering – long run time, a few models failed to converge

Lessons learned: proactive steps

- More computing memory, server capacity
- Advanced coding for data cleaning, code review
- Test runs
  - Troubleshot issues with model convergence

Data Extraction and Quality:
what I know now that I wish I knew then

Corita R. Grudzen, MD, MSHS, FACEP
Division Head, Supportive and Acute Care Services
Fern Grayer Chair in Oncology Care and Patient Experience
Director, Center for Cancer Care Innovation
Memorial Sloan Kettering Cancer Center
Professor of Emergency Medicine
Weill Cornell Medical College

Memorial Sloan Kettering
Cancer Center

Study Phase

- Planning
  - Pre-randomization must dos!
- Implementation
  - Monitoring of feasibility, acceptability, usability or fidelity
- Analysis
  - Advantages and disadvantages of push versus pull

Planning phase

Critical to examine BEFORE randomization:

- Assess infrastructure (*e.g.,* brick and mortar, staffing, software or IT resources)
- Check interoperability of data systems or electronic health record
- Evaluate data sharing capability and data use agreements
- Assess quality of outcome data
- “Willingness” or “get it done” attitude
- ….
- ….
- ….
- Adequate volume of eligible patients

Planning phase examples

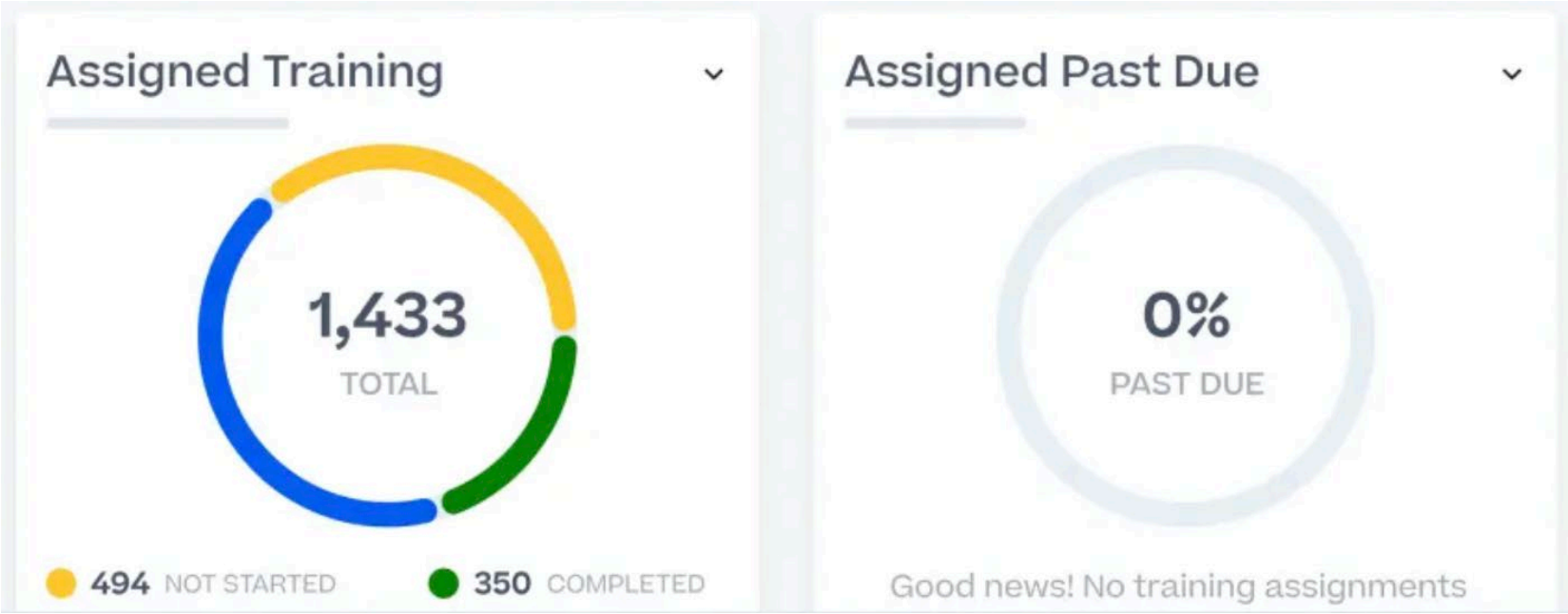
- Human resources
  - Outpatient specialty palliative care practice
  - ED social worker
- Electronic health record (*e.g.,* Epic)
  - Transition mid-project
- Data sharing
  - Test ability and willingness to obtain quality data, especially if PHI
- Inability to identify sites (free-standing EDs) in CMS data
- “Willingness” or “get it done” attitude can overcome almost anything!

Implementation

- Training data
- Feasibility (volume of patients, procedures, visits)
- Acceptability (survey completion)
- Fidelity to core function
- Usability of electronic tools

Implementation examples

- Learning Management System



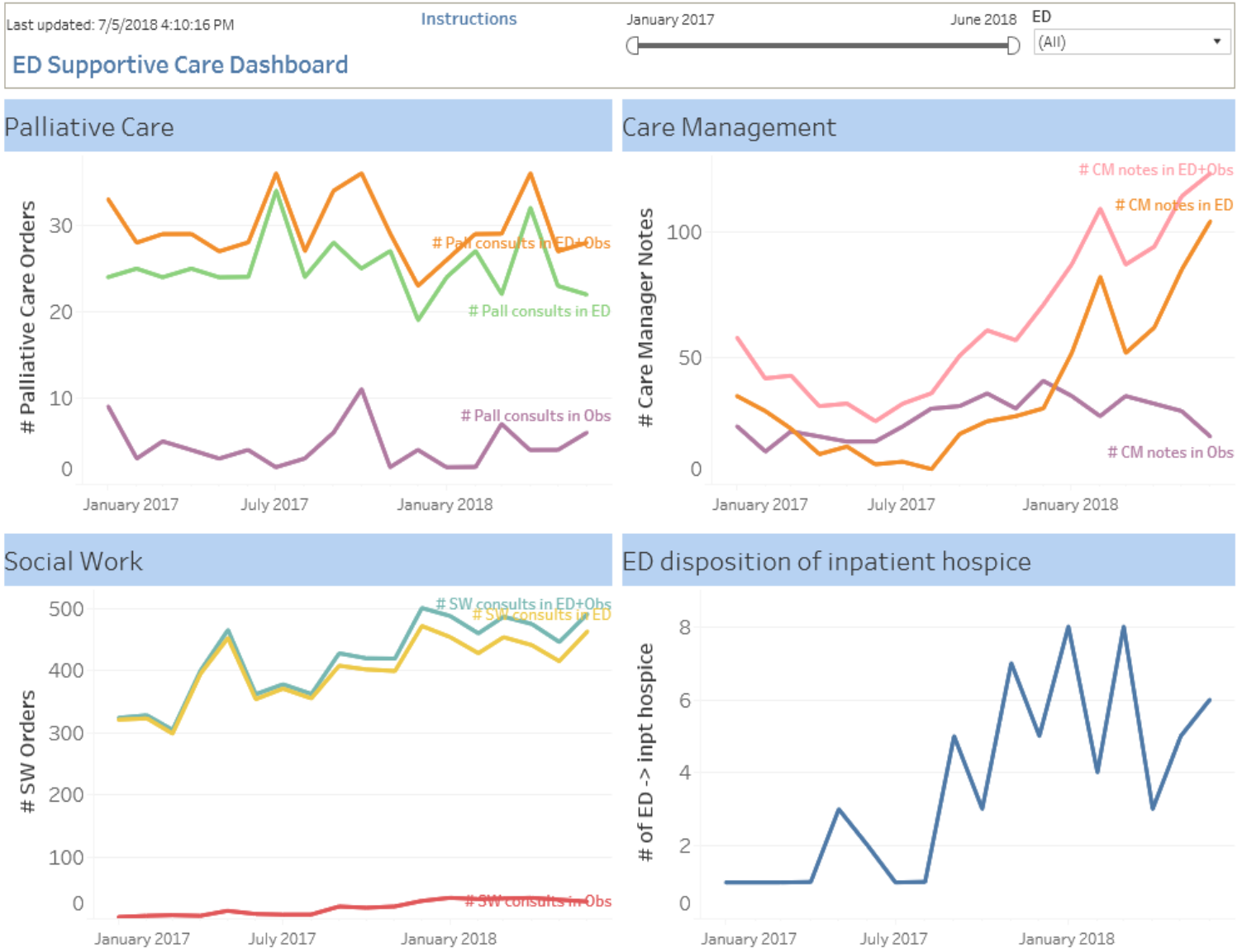
Implementation examples

- Survey completion rates
  - RECCap, Qualtrix

Record ID	Enrollment			Baseline data		Post data	
	Records	Sites	Student Survey	Test 1	Test 2	Test 1	Test 2
1001	●	●	●	●	●	●	○
1002	●	●	●	●	●	●	○
1004	●	●	●	●	●	●	○
1005	●	●	●	○	○	○	○
1009	●	●	●	○	●	●	○
1010	●	●	●	○	●	●	○
1011	●	●	●	●	○	●	○
1012	●	●	●	●	○	●	○
1013	●	●	●	●	●	●	○
1014	●	●	●	○	●	●	●
1015	●	●	●	●	●	●	●
1017	●	●	●	●	●	●	●
1020	●	●	●	●	●	●	●
1021	●	●	●	●	●	○	○
1022	●	●	●	●	●	○	○
1025	●	●	●	●	●	○	○

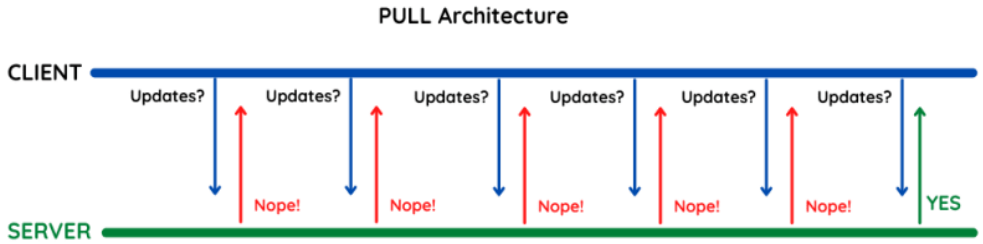
Implementation examples

- Clinical Decision Support
  - Process metrics

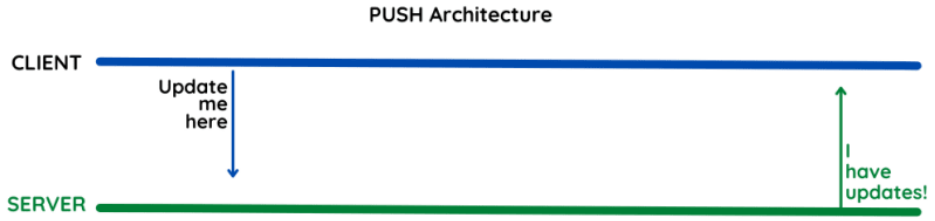


Analysis

- Pull architecture
  - Driven by a request
  - Requires more human resources
  - Pdf or excel spreadsheet



- Push architecture
  - Driven by an event and automatically pushed
  - Requires more technical expertise and lead time
  - Code provided to other sites



Memorial Sloan Kettering
Cancer Center

Questions