

Lumbar Imaging With Reporting Of Epidemiology (LIRE)—Protocol for a pragmatic cluster randomized trial☆



Jeffrey G. Jarvik^{a,b,c,h,*}, Bryan A. Comstock^{d,g}, Kathryn T. James^{a,h}, Andrew L. Avinsⁿ, Brian W. Bresnahan^{a,h}, Richard A. Deyo^{j,k,l,m}, Patrick H. Luetmer^p, Janna L. Friedly^{f,h}, Eric N. Meier^{d,g}, Daniel C. Cherkin^q, Laura S. Gold^h, Sean D. Rundell^{f,h}, Safwan S. Halabi^o, David F. Kallmes^p, Katherine W. Tan^{d,g}, Judith A. Turner^{e,f,h}, Larry G. Kessler^c, Danielle C. Lavalleyⁱ, Kari A. Stephens^e, Patrick J. Heagerty^{d,g}

^a Department of Radiology, University of Washington, USA

^b Department of Neurological Surgery, University of Washington, USA

^c Department of Health Services, University of Washington, USA

^d Department of Biostatistics, University of Washington, USA

^e Department of Psychiatry and Behavioral Sciences, University of Washington, USA

^f Department of Rehabilitation Medicine, University of Washington, USA

^g Center for Biomedical Statistics, University of Washington, USA

^h Comparative Effectiveness, Cost and Outcomes Research Center, University of Washington, USA

ⁱ Surgical Outcomes Research Center, University of Washington, Seattle, USA

^j Department of Family Medicine, Oregon Health Sciences University, Portland, OR, USA

^k Department of Internal Medicine, Oregon Health Sciences University, Portland, OR, USA

^l Department of Public Health and Preventive Medicine, Oregon Health Sciences University, Portland, OR, USA

^m Oregon Institute for Occupational Health Sciences, Oregon Health Sciences University, Portland, OR, USA

ⁿ Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA

^o Department of Radiology, Henry Ford Hospital, Detroit, MI, USA

^p Department of Radiology, Mayo Clinic, Rochester, MN, USA

^q Group Health Research Institute, Seattle, WA, USA

ARTICLE INFO

Article history:

Received 10 August 2015

Received in revised form 9 October 2015

Accepted 11 October 2015

Available online 19 October 2015

Keywords:

Pragmatic randomized trial

Cluster randomized trial

Back pain

Spine imaging

Lumbar imaging

Stepped wedge design

ABSTRACT

Background: Diagnostic imaging is often the first step in evaluating patients with back pain and likely functions as a “gateway” to a subsequent cascade of interventions. However, lumbar spine imaging frequently reveals incidental findings among normal, pain-free individuals suggesting that treatment of these “abnormalities” may not be warranted. Our prior work suggested that inserting the prevalence of imaging findings in patients without back pain into spine imaging reports may reduce subsequent interventions. We are now conducting a pragmatic cluster randomized clinical trial to test the hypothesis that inserting this prevalence data into lumbar spine imaging reports for studies ordered by primary care providers will reduce subsequent spine-related interventions. **Methods/design:** We are using a stepped wedge design that sequentially randomizes 100 primary care clinics at four health systems to receive either standard lumbar spine imaging reports, or reports containing prevalence data for common imaging findings in patients without back pain. We capture all outcomes passively through the electronic medical record. Our primary outcome is spine-related intervention intensity based on Relative Value Units (RVUs) during the following year. Secondary outcomes include subsequent prescriptions for opioid analgesics and cross-sectional lumbar spine re-imaging.

Discussion: If our study shows that adding prevalence data to spine imaging reports decreases subsequent back-related RVUs, this intervention could be easily generalized and applied to other kinds of testing, as well as other conditions where incidental findings may be common. Our study also serves as a model for cluster randomized trials that are minimal risk and highly pragmatic.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Diagnostic imaging is often an early step in the work-up of back pain and a likely gateway to subsequent interventions. Unfortunately, these studies reveal incidental anatomic spine findings among many normal, pain-free individuals. [1–3] Such findings can be alarming to clinicians

☆ Funding source: NIH 1UH2AT007766-01 and 4UH3AR066795-02. Trial Registration: Clinicaltrials.gov NCT02015455.

* Corresponding author at: Department of Radiology, University of Washington, Box 359728, 325 Ninth Ave., Seattle, WA 98104-2499, USA.

E-mail address: jarvik@uw.edu (J.G. Jarvik).

and patients alike, and may prompt unnecessary additional tests and treatments. [4–6].

Because many spine findings seen on spine radiography may be incidental and not responsible for the patient's symptoms, Roland and van Tulder proposed adding statements to spine imaging reports describing the prevalence of various degenerative findings among patients without back pain. [7] In a prior pilot study, our group found that primary care patients undergoing lumbar spine imaging were less likely to receive a variety of subsequent diagnostic and therapeutic interventions if their imaging reports contained information describing the prevalence of common imaging findings among patients without back pain. However, this initial study was small and retrospective. More definitive conclusions regarding the impact of adding benchmark prevalence information to radiology reports requires a large, prospective randomized controlled trial, leading to our designing the Lumbar Imaging with Reporting of Epidemiology (LIRE) trial.

Our main hypothesis is that for patients of primary care providers, inserting age- and imaging modality-appropriate benchmark prevalence data into lumbar spine imaging reports will reduce overall spine-related healthcare utilization and testing as measured by our primary outcome, spine-related relative value units (RVUs), including magnetic resonance imaging (MRI) and computed tomography (CT), as well as subsequent therapeutic interventions, such as spinal injections, and spine surgeries. An important secondary hypothesis is that the intervention will reduce subsequent opioid prescriptions, as suggested by our pilot work. [8] Because advanced imaging modalities (MRI and CT) are more sensitive than radiographs for detecting a variety of incidental findings, [9,10] a third hypothesis is that there will be a greater reduction in subsequent spine-related diagnostic and therapeutic interventions for MRI and CT compared with radiography. Our fourth hypothesis is that there will be a greater impact for findings such as disc bulges that are likely less clinically important than other findings, such as disc extrusions.

2. Materials and methods

2.1. Overview

We aim to conduct a pragmatic, cluster randomized trial, assigning primary care clinics at 4 large health systems to receive either standard lumbar spine imaging reports or reports containing age- and modality-appropriate epidemiological benchmarks for common imaging findings. We will use a stepped wedge design that randomizes participating clinics to initiate the intervention sequentially on one of five pre-specified calendar times: April 1, 2014, October 1, 2014, April 1, 2015, October 1, 2015, and April 1, 2016. The stepped wedge design allows for both between-clinic cross-sectional and within-clinic before/after longitudinal comparisons of the intervention, while assuring that all clinics eventually receive the intervention. Our primary outcome is a summary measure of back-related intervention intensity, spine-related relative value units (RVUs). We based our measure of spine-related RVUs on work done by Martin and colleagues and have devoted substantial effort toward refining the measure for use with data that we can retrieve from the participating health systems. [4,11].

2.1.1. Participating centers

We are enrolling patients at four integrated health care systems: Kaiser Permanente, Northern California (KPNC) Oakland, CA; Henry Ford Health System (HFHS) Detroit, MI; Group Health Cooperative (GHC), Seattle, WA; and Mayo Clinic Health System (MCHS), Rochester, MN. These sites contribute both geographic and demographic diversity and have comprehensive electronic medical record (EMR) systems, allowing capture of health care utilization data.

The University of Washington's Comparative Effectiveness, Cost and Outcomes Research Center (CECORC) and Center for Biomedical Statistics (CBS) serve as the Data Coordinating Center (DCC) for LIRE.

A collaborator at Oregon Health and Science University (RAD) is also part of the DCC.

2.2. Study design

2.2.1. Eligibility criteria-clinics

We define a "clinic" as primary care if a majority of its practitioners provide adult primary care (general internal medicine, family practice, and associated mid-level providers). The criteria for clinic eligibility are that the health care providers are a distinct, readily identifiable group that has at least a subgroup of primary care providers who do not practice at another clinic that will also be part of the trial. The requirement that providers be based primarily at one clinic is to minimize cross-contamination (having the epidemiological benchmarks at one clinic influence another clinic that is not yet receiving the epidemiological benchmarks).

2.2.2. Eligibility criteria-patients

Patients are eligible for the LIRE study if a primary care provider orders a diagnostic imaging test of the lumbar spine between October 1, 2013, and September 30, 2016. We include all adult patients (≥ 18 years old as of the date of the imaging test) who undergo plain film, CT, or MR imaging of the lumbar spine, identified from Current Procedural Terminology (CPT) codes [12] (see Table 1) by primary care providers.

All patients receiving eligible lumbar spine imaging studies at participating clinics are part of the trial, unless the patient has explicitly signed a declaration opting out of all research studies.

2.2.3. Patient identification

The health information system automatically identifies when a practitioner from a participating clinic orders an eligible lumbar spine imaging study. We accomplish this either through the radiology information system (RIS) or EMR, depending on the site.

2.2.4. Consent

The Institutional Review Boards (IRBs) at KPNC and UW ceded authority to the Group Health Research Institute (GHRI) IRB, which is the IRB of record for the overall study. The IRBs at HFHS and Mayo retained review of the protocol. Each reviewing IRB approved the trial procedures. All participating IRBs agreed that our study is minimal risk and granted waivers of both consent and Health Insurance Portability and Accountability Act (HIPAA) authorization. One site required the investigators to inform primary care providers at the end of the study that we conducted a randomized evaluation at their site.

2.2.5. Randomization

We randomly assigned all clinics at each site to receive the intervention at one of five fixed calendar times: April 2014, October 2014, April 2015, October 2015, and April 2016 (Fig. 1).

Within each recruitment site, we sorted clinics by number of primary care providers (PCPs) into tertiles (small, medium, large clinics).

Table 1
Eligible lumbar spine imaging procedures by CPT codes.

CPT code	Description
72080	Spine, thoracolumbar, 2 views
72100	Spine, lumbosacral; 2 or 3 views
72110	Spine, lumbosacral; minimum of 4 views
72114	Spine, lumbosacral; complete, with bending views
72131	CT, lumbar spine without contrast
72132	CT, lumbar spine with contrast
72133	CT, lumbar spine without and with contrast
72148	MRI, lumbar spine without contrast
72149	MRI, lumbar spine with contrast
72158	MRI, lumbar spine without and with contrast

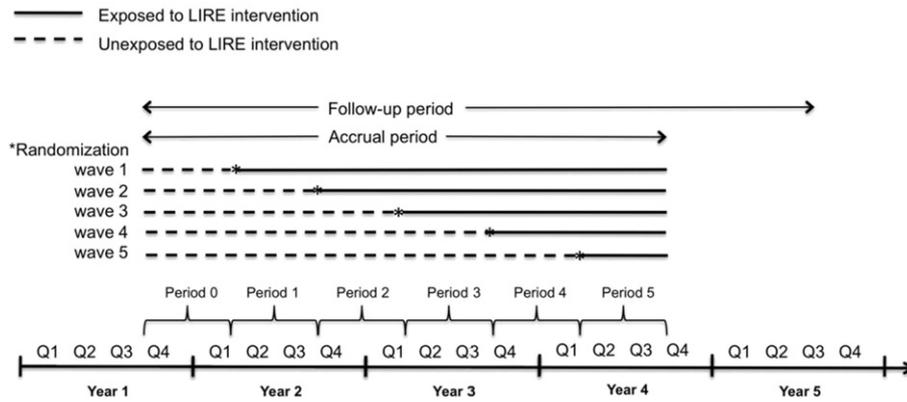


Fig. 1. Randomization scheme.

From each tertile we randomly selected clinics using urn-based randomization (without replacement) stratified by site and clinic size such that clinics of small, medium, and large size are equally represented in each randomization wave. We use site-specific definitions for the size of the clinic with the goal of having balance of clinic size across randomization waves within each site. Because clinic size varied substantially both within and between health systems, balancing randomization on clinic size ensures comparable time in the control and intervention periods for all clinic size strata. Table 2 displays the site-specific strata definitions and size.

Due to providers practicing at multiple clinics, we combined some clinics into single randomization units to reduce cross-contamination, resulting in a total of 100 randomization units. We initially identified 1627 PCPs as units of observation within those randomization units, realizing that the number of providers is in constant flux due to new hires, retirement, etc. Additionally, because of uncertainty in classifying providers as PCPs at one of our sites (KPNC) we expect that the total number of participating PCPs will rise to between 2400 and 4400.

2.2.6. Intervention

Using an automated approach through either the RIS or EMR, we insert age- and modality-appropriate epidemiological benchmark information [3] automatically into the lumbar spine imaging reports of PCPs who work in intervention clinics. The exact method of insertion varies by site.

At GHC, we use the EMR (Epic) to insert the intervention text after the radiologist has finalized the radiology report. Similarly at Mayo, we insert the intervention text after the radiologist has finalized the report, using a system that interfaces with the RIS and the EMR (Cerner). At KPNC and HFHS, the RIS inserts the intervention at the time that the radiologist dictates the report. At KPNC, this allows the radiologist to review the text and potentially modify or remove it. If they do either, they must insert a text string before they are able to finalize the report that allows us to track these modifications.

We based the intervention text (see Appendix A) on text that our group had previously used [7,8], with modifications that included

updates from a systematic review of the prevalence of findings in patients without back pain as well as input from patients to make the text more easily understood by non-medical readers. Primary care providers who work in control clinics receive their usual imaging report.

2.2.7. Data collection methods

We collect all baseline and follow-up data from the electronic information systems which, depending on the site, include EMR and administrative data systems. We use the PopMedNet software application that enables simple creation, operation, and governance of distributed health data networks. [13,14] This system provides a secure, mutually governed and auditable model for transporting data between the sites and the DCC. Each LIRE partner site is also a participant in the LIRE PopMedNet network and has a named and authenticated set of users.

For EMR queries, we perform two types of data extraction: 1) index files consisting of all patients receiving an eligible lumbar imaging study at participating clinics, with data limited to date and type of imaging procedure, clinic identifier, patient demographics, imaging report text; and 2) comprehensive data extraction that includes outcome and safety variables (described below). We perform an index data file query 2–3 weeks after the start of each randomization wave to verify that the intervention is being deployed appropriately. For all of the imaging text reports included in each index data set, we use text matching to verify that the intervention text was properly inserted into the reports based on clinic, imaging modality, and patient age.

One year following the first randomization wave and then every six months thereafter, we perform a comprehensive data query for both safety monitoring and outcomes assessment. These files are compatible with the Virtual Data Warehouse schema, and we will implement them at each of the HMO Research Network (HMORN) sites (KPNC, HFHS, and GHC) using common SAS software code. The Virtual Data Warehouse is a protocol for obtaining common data elements from health systems that are part of the HMORN. [15] This facilitates data extraction as well as comparisons between these health systems. MCHS, which is not part of the HMORN, developed a local mapping of the common data dictionary to provide semantically identical data index files and summary data files on the same schedule. Each site retains full lists of eligible patients and will be responsible for addressing patient withdrawals, additions, and deaths. Sites submit all index and comprehensive outcome files as Limited Data Sets (de-identified except for dates of service) to the DCC via PopMedNet. Sites provide unique, coded study identifiers for each patient so that index image data can be linked to comprehensive outcomes and so that patients with more than one lumbar spine image during the study period are only included in the study once (images subsequent to the index image become part of the outcome data). The DCC maintains the data dictionary for the duration of the project.

Table 2 Within-site stratified randomization schedule of clinics by number of PCPs.

Recruitment site	Clinics/units of randomization (# of PCPs)	PCP strata size boundaries (# clinics)		
		Small	Medium	Large
Site #1	19 (245)	5 to 10 (7)	11 to 15 (7)	16 to 41 (5)
Site #2	26 (230)	3 to 6 (9)	7 to 9 (9)	10 to 24 (8)
Site #3	21 (814)	7 to 29 (7)	33 to 39 (7)	43 to 106 (7)
Site #4	34 (338)	2 to 4 (12)	5 to 9 (10)	11 to 34 (12)
Total	100 (1627)			

2.2.8. Pre-intervention data collection

To determine baseline pre-intervention RVUs and outcomes prior to randomization, we included all patients at participating clinics who received eligible lumbar spine imaging studies between October 1, 2013 and March 31, 2014. These data comprised our baseline period. Additionally, we are collecting diagnosis and utilization data for patients 12 months prior to baseline (potentially as early as October 1, 2012 for a patient who received index imaging on October 1, 2013) to characterize the assembled cohort at a patient-level.

2.2.9. Data collection schedule

We will capture EMR data on patients for a minimum of one year after the index imaging examination. Two-thirds of patients will have two years of EMR-based follow-up data due to the staggered implementation of the intervention. Biannually, we will collect comprehensive EMR data for patients who reached their one and two-year follow-up date in the previous six months.

2.2.10. Primary and secondary outcomes

We will collect all outcomes passively through the EMR (Table 3). The primary outcome is the one-year summary of lumbar spine-specific RVUs, a composite measure of spine interventions that combines the overall intensity of resource utilization for back pain care into a single metric.

We developed a summary spine-specific RVU metric using EMR data from a cohort of 5239 patients with back pain as part of the Back pain Outcomes using Longitudinal Data (BOLD) project. [16] We developed a data dictionary and code to abstract EMR data across three health systems, two of which are LIRE sites. For each BOLD participant, we obtained a comprehensive list of procedures (CPT codes), diagnoses and provider visits (International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes), [17] and inpatient hospitalizations from the EMR. Using the Medicare Physician Fee Schedule (<http://www.cms.gov/>), we generated and tested a mapping algorithm to assign RVUs to over 10,000 unique CPT codes. A sample of RVUs from the 2012 CMS file is shown in Table 4. We will not capture care provided to patients outside the partner health systems that is not included in their EMRs.

To obtain a spine-related summary RVU from CPT and ICD-9-CM codes, we used an existing validated algorithm. [11,18,19] Aggregating across CPT codes identified by this algorithm through one year after the index diagnostic imaging test yields the total spine-specific RVUs. Data extraction from the EMRs will likely be impacted by the conversion to ICD10. Because ICD10 is more granular than ICD9, it is possible to develop a cross-walk algorithm that accounts for all of the data elements that we identify using ICD9 from an ICD10 dataset. Since each health system EMR has idiosyncratic elements, we plan to develop this crosswalk for each health system.

Table 3
Outcomes data from EMR.

Domain	Specific Element
Hospitalizations (all inpatient stays)	<ul style="list-style-type: none"> Duration (days) Diagnosis related group (DRG) Associated Current Procedural Terminology (CPT) and International Classification of Diseases-9 (ICD-9) codes
Outpatient visits	<ul style="list-style-type: none"> Clinic type Associated ICD-9 codes
Pharmacy data (prescribed medications)	<ul style="list-style-type: none"> National Drug Code (NDC) Drug name, dose, quantity, days' supply
Procedures (inpatient and outpatient, including subsequent imaging)	<ul style="list-style-type: none"> Procedure type and indication Associated CPT and ICD-9 codes
Safety Data	<ul style="list-style-type: none"> ER visits (within 90 days of index image) Death (within 6 months of index image)

Table 4
Selected spine-related CPT codes and associated RVUs.

CPT code	Description	RVU
72100	X-ray exam of lower spine – 2 or 3 views	1.07
97001	Physical therapy evaluation	2.18
99214	Detailed office visit	2.26
99284	Emergency department visit – high intensity	3.37
64483	Epidural injection for lumbar spinal stenosis	3.37
72131	CT lumbar spine without contrast	6.27
72148	MRI Lumbar Spine without contrast	11.31
63047	Removal of spinal lamina (laminectomy)	32.89
22804	Fusion of the spine	71.60

We will also derive important secondary outcomes from EMR and administrative data. These include: a longer-term back-specific RVU summary metric reflecting back-related care up to two years following the index imaging study date; an indicator of opioid prescribing after the index imaging study; subsequent advanced imaging (number of MRI or CT studies) within 90 days and 12 months after the index imaging study; spine injections, spine surgery and other back-related medical costs over the 2 years.

2.2.11. Imaging finding abstraction

We will use an approach that combines statistical machine learning with rules based natural language processing (NLP), to abstract the necessary data from anonymized radiology text imaging reports. The health systems will export the imaging reports from their EMR and we will identify common imaging findings that are likely less clinically important (e.g., disc bulge, disc space narrowing, annular fissure) versus more important (central canal stenosis, nerve root compression, disc extrusion) (Table 5). [1,20] We will also identify less common but potentially clinically important findings such as possible tumor or infection.

For the machine learning approach, we plan to use the NLTK library in Python. [21] We then plan to create a “feature set” of variables derived from report text that optimizes prediction. Features included will likely include basic N-grams (unigrams, bigrams, trigrams), sections of the report, and type of imaging examination.

For the rule-based portion, we plan to code regular expressions (REGEX) based on a list of synonyms for imaging findings developed by our study content experts (JGJ, SDR, RAD). We also plan to account for negation by adapting the ConText algorithm [22]. We will evaluate our prediction accuracies using a random subset of annotated reports. Based on a pilot investigation, agreement among the three annotators was substantial (kappa ranging from 0.65–1) for various imaging findings.

We will use these data to examine whether there is any differential effect of the intervention according to the results in the imaging report.

Table 5
Imaging findings likely clinically important vs. likely clinically unimportant.

Likely clinically important	Likely clinically unimportant
<ul style="list-style-type: none"> Moderate or severe stenosis* Disc extrusion Nerve root displacement or compression Endplate edema (Type 1 endplate change) Grade 2 or higher listhesis 	<ul style="list-style-type: none"> Annular Fissure Disc height loss Mild stenosis (central, lateral recess or foraminal) Nerve root contact without displacement/compression Grade 1 listhesis Disc desiccation Disc bulge Disc protrusion Facet degeneration (any severity)

* Central, lateral recess or foraminal.

2.2.12. Data quality control (QC)

As biannual EMR data accumulate, we will use established algorithms to ensure data quality and validation. We will focus evaluation of EMR data quality primarily on patterns of completeness and substitution (use of one code for another) in CPT codes (inpatient and outpatient), ICD-9-CM diagnosis codes, and NDC pharmacy records. For QC of patient-level EMR data, we will check that the data are of the appropriate type and format specified by the data dictionary. For system-to-system and wave-to-wave QC evaluation of EMR data, we will generate diagnostic plots comparing rates of CPT code endorsement to look for unusual patterns of utilization or use of specific CPT codes. [23] Similarly, we will use site-by-site cross-tabulation and diagnostic plots to examine patterns in prescribed medications by NDC code. In an additional QC step, we will cross-tabulate the top 100 CPT codes by frequency and RVUs for each site in order to detect unusual patterns of missing codes or code substitution. We will document and resolve any issues that we discover through direct discussion with programming staff and site principal investigators.

2.3. Analytic approach

To evaluate the impact of inserting prevalence data into an imaging report, we will use longitudinal regression methods such as linear mixed effects models (LMMs) or generalized linear mixed models (GLMMs) with robust standard errors for all primary and secondary outcome measures. Mixed models provide an efficient method for analysis of longitudinal or multilevel data and will be the basis of our primary analysis approach. In secondary analyses we will use generalized estimating equations (GEE), adopting simple exchangeable correlation models at the clinic level to determine whether conclusions appear sensitive to model specification.

We will use a time-varying intervention status indicator $Status_{kt}$ ($0 = \text{control}$, $1 = \text{intervention}$, for clinic k at time t) for the primary longitudinal model for back pain specific RVUs. We will adopt the functional form given below for the specific regression model, with fixed effects for time (linear), age (18–39, 40–59, 60+, using 2 dummy variables), imaging modality type (plain film or MRI, using dummy variables), clinic size (small, medium, large, using 2 dummy variables), and site (GHC, HFHS, KPNC, MCHS, using 3 dummy variables) in addition to random effects for provider, clinic, and intervention status:

$$Y_{ijk} = \beta_0 + \beta_1 \cdot \text{Time}_t + \beta_2^T \cdot \text{Age}_{ijk} + \beta_3^T \cdot \text{Modality}_{ijk} + \beta_4^T \cdot \text{Size}_k + \beta_5^T \cdot \text{Site}_k + \lambda_0 \cdot \text{Status}_{kt} + \text{mean model}$$

$$b_{k,0} + b_{k,1} \cdot \text{Status}_{kt} + \text{clinic random effects}$$

$$a_{jk,0} + e_{ijk} \text{ provider random effects and subject errors.}$$

We will collect the outcome measure Y_{ijk} on patient i ($i = 1, 2, \dots, n_j$) under primary care provider j ($j = 1, 2, \dots, n_k$) enrolled in time period t ($t = 0, 1, 2, \dots, 5$) in order to evaluate the overall effect of the intervention at the level of the clinic k ($k = 1, 2, \dots, 100$). The primary parameter of interest, λ_0 , represents the average effect of the intervention adjusting for temporal trends (Time_t), clinic characteristics (Site_k , Size_k), and individual covariates (Age_{ijk} , Modality_{ijk}).

Our second hypothesis is that the intervention will reduce subsequent opioid prescriptions. Our pilot data suggested this effect and it could be mediated by the intervention reassuring providers and patients regarding the benignity of imaging findings. Each clinical site data system utilizes the national drug code or similar classification [24] that allows us to determine the morphine equivalent dose (MED) for each opioid prescription. [25] We will then calculate the total MED prescribed per patient within 90 days and 12 months of the index image. We will then use the same approach for the analysis of MED as we used for back-specific RVUs, using a time-varying intervention

status indicator $Status_{kt}$ ($0 = \text{control}$, $1 = \text{intervention}$, for clinic k at time t) for the primary longitudinal model for MED prescriptions.

Our third hypothesis is that there will be a differential effect of the intervention according to imaging modality: there will be a greater reduction in subsequent spine-related diagnostic and therapeutic interventions for MRI compared with radiography. In order to test this hypothesis, we will analyze patient-level data according to the appropriate LMM or GLMM given above, but including the interactions between Modality_{ijk} indicators (modeled using an indicator variable coding MRI, with plain film as the reference) and Status_{kt} . A test of the interaction terms (2° -of-freedom Wald test) will be used to test the null hypothesis that the effect of the intervention does not vary according to the imaging modality.

Our fourth hypothesis is that there will be a greater impact for findings that are likely less clinically important than findings that are likely more clinically important. We will use an additional variable, $\text{ImageFinding}_{ijk}$, that takes the value 1 if an important image finding is present, and 0 otherwise (see Table 5). We will test the null hypothesis that the interaction between $\text{ImageFinding}_{ijk}$ and Status_{kt} is zero using a Wald test.

2.3.1. Power for primary outcome

We randomized 100 clinics with approximately 1700 PCPs. We calculated statistical power for the primary outcome measure, spine-specific RVUs, using data from the BOLD study to inform key design parameter estimates. From BOLD, we identified 639 patients at KPNC and HFHS who had a qualifying lumbar image within six weeks of a PCP visit, the majority (74%) of which occurred within seven days. As expected, patient-level RVUs were positively skewed and we will therefore use an approximately normalizing transformation of $\log(\text{RVU} + 1)$ but make interpretations regarding effect size on the original RVU scale.

With log-transformed BOLD Registry RVU data, we fit a linear mixed effects model adjusting for image type (advanced vs. plain film) and study recruitment site and estimated the variance components for clinic (0.026) and the residual error term (1.230). The observed intra-class correlation coefficient (ICC) across clinics was 0.013 (95% CI: 0.000 to 0.046). In this subset of BOLD data, the number of PCPs with multiple patients was too few to inform the PCP-level variance component and it is therefore conservatively included in the error term variance for power simulations.

We considered fixed numbers of clinics and PCPs for each simulation and we assumed that each provider would provide data for all study time periods. For a range of potential RVU effect sizes, we generated 1000 simulated data sets and performed mixed model estimation with each data set. For the primary outcome measure of back-specific RVUs, the study has 89% power to detect reductions of 5.0% or larger in the median back-specific RVU. For a patient receiving a lumbar image, a 5% average reduction in spine-specific RVU translates into approximately one fewer lumbar CT scan (for example) on average compared to a patient unexposed to the LIRE intervention.

2.3.2. Data safety monitoring plan

Two external safety officers will monitor safety endpoints (Emergency Room (ER) visits within 90 days and deaths within six months of index imaging) every six months for the duration of the study. Safety officers will utilize absolute relative risk ratio monitoring thresholds of 1.15 and 1.10 for comparing 90-day ER visit and death rates by intervention group, with adjustment for pre-intervention rates at each site and patient-specific Quan comorbidity index. [26].

3. Discussion

The imaging of patients with back pain has the potential for both benefit and harm. Imaging can identify conditions such as central spinal stenosis that in the appropriate clinical setting has a good evidence base for the benefits of certain treatments such as surgery. [27] However,

because of the high prevalence of a variety of imaging findings in patients without back pain, imaging may lead to a cascade of needless interventions that can be both harmful and expensive. We describe a study that tests a strategy to mitigate the problematic consequences of potentially misleading lumbar spine imaging findings by including benchmark prevalence data in the routine imaging report.

We adopted the concept of adding epidemiologic benchmarks from Roland and van Tulder, who proposed adding statements to plain film imaging reports describing the prevalence of different degenerative findings in patients without back pain. [7] A small, retrospective, pilot study published by our group suggested that patients who had the benchmark epidemiological information included in their radiology report were about three and a half times less likely to get a narcotic prescription compared with patients who did not get this information in their imaging report. [8] This was statistically significant after controlling for a variety of factors (age, gender, severity of condition) (OR = 0.29, $P = 0.01$). Not statistically significant but also reduced in the “statement” group were subsequent high-cost imaging examinations (CT and MRI) (OR = .22) and physical therapy (PT) referrals (OR = .55, $P = 0.06$).

Why might this simple, inexpensive approach be effective? We surmised that the benchmark information would reassure both patients and physicians and result in fewer downstream interventions, including both diagnostic (cross-sectional imaging) and treatment (PT, opiates, surgeries) interventions.

LIRE is a pragmatic cluster randomized trial of a minimal-risk intervention that we believe can serve as a model for future pragmatic trials. We designed LIRE to be as pragmatic as possible. Thorpe et al. [28] described a tool that they named the pragmatic–explanatory continuum indicator summary (PRECIS) to determine where on the spectrum of explanatory to pragmatic a trial falls with respect to 10 domains, and it has recently been modified to the PRECIS-2 tool by Loudon et al. [29]. Fig. 2 is the PRECIS-2 diagram of LIRE, demonstrating the highly pragmatic nature of the study design. [30].

LIRE is 1 of 7 pilot pragmatic randomized trials that are part of the initial round of funding by the NIH Health Care Systems Collaboratory. [31] LIRE and the other pilot pragmatic trials are a first step toward that goal.

One potential limitation of LIRE is that the intervention may lead PCPs to devalue the importance of all imaging findings and thus under-treat conditions that could have benefitted from treatment. However, given the climate of over-diagnosis and over-treatment of back pain in the U.S., this is less likely to occur than in other settings. Another limitation is that, since we are not collecting patient reported outcomes (PROs), we cannot directly comment on the effect of the intervention on potentially important patient outcomes such as functional status and pain. We are also not able to collect information on work loss or disability. Nor can we comment on the effectiveness of the intervention on potentially important subgroups such as patients

with acute vs. chronic back pain or those who satisfy the National Quality Forum’s criteria for imaging. We made this decision not to collect patient reported data deliberately, conscious of the fact that collecting PROs would make the trial less pragmatic. Given our projected sample size of over 160,000 patients, collecting PROs would not have been practical.

Because of the large size of our project, abstraction of the radiology reports is a challenge. We plan to use natural language processing as a tool to help us with this abstraction task. Natural language processing is reasonably accurate at identifying clinical findings from radiology reports. Chapman et al. reported sensitivities of 86–92%, specificities of 78–91%, and positive predictive values (PPV) of 72–85% for NLP in X-ray reports [32]. Despite this reported success as well as our promising pilot experience, once we start applying NLP to the larger dataset we may discover obstacles that will prompt us to revise our approach.

If the current trial shows that adding prevalence data to spine imaging reports decreases subsequent back-related RVUs, the method is likely to be generalizable to other conditions and to other kinds of testing (e.g., other imaging procedures such as chest CT, abdominal CT) where incidental findings may be common. [33] Because the cost of the intervention itself is minimal, the potential cost-effectiveness of this intervention, if successful, is enormous. If it reduces unnecessary treatment, the intervention could prevent avoidable complications, decreasing costs and improving patient outcomes.

Our project, the Lumbar Imaging with Reporting of Epidemiology (LIRE) study, also serves as a model for a cluster randomized trial that is minimal risk and highly pragmatic. If successful, our use of the stepped wedge design on a large scale could become a desirable model for future studies and represent an important strategy for health services researchers.

4. Conclusion

The long-term public health significance of the LIRE study is that, if effective, our simple, inexpensive intervention has the potential to reduce unnecessary, inappropriate, and costly care not only for back pain, but also for a wide range of other conditions, since it could easily be applied to other diagnostic tests (e.g., other imaging tests, laboratory tests, genetic testing). If our study demonstrates that the intervention decreases subsequent health-resource use, adding epidemiologic benchmarks to diagnostic test reporting could become the dominant paradigm for communicating all diagnostic information, resulting in substantially more appropriate clinical care for minimal implementation costs.

Abbreviations

ANCOVA	analysis of co-variance
BOLD	Back pain Outcomes using Longitudinal Data
CPT	Current Procedural Terminology
DCC	data coordinating center
EMR	electronic medical record
ER	emergency room
GEEs	generalized estimating equations
GLMMs	generalized linear mixed models
HIPAA	Health Insurance Portability and Accountability Act
ICD-9-CM	International Classification of Diseases, Ninth Revision, Clinical Modification
LMMS	linear mixed effects models
NDC	National Drug Codes
NLP	natural language processing
QCC	quality control
RIS	radiology information system

Competing interests

Dr. Jarvik has the following potential conflicts of interest, although they do not relate directly to the subject of this manuscript, he lists

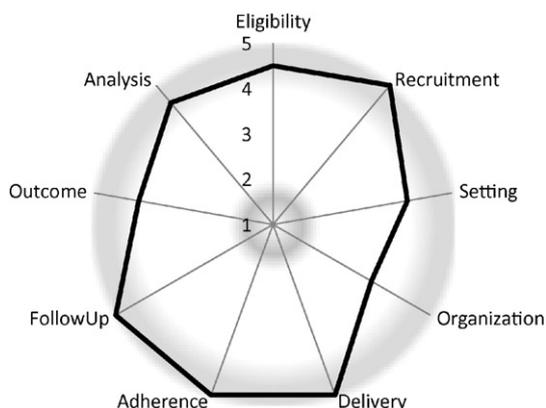


Fig. 2. Pragmatic-explanatory continuum indicator summary (PRECIS).

them in the spirit of full disclosure. He has served on the Comparative Effectiveness Advisory Board for GE Healthcare through October 2012. He is a co-founder and stockholder of PhysioSonics, a high intensity focused ultrasound company, and receives royalties for intellectual property. He serves as a consulting medical editor for Google as well as a consultant for HealthHelp, a radiology benefits management company. Finally, he is a co-Editor of Evidence-based Neuroradiology published by Springer Publishing.

Authors' contributions

JGJ, BAC, RAD, PJH developed the original concept of the study and JGJ, BAC, ALA, BWB, RAD, JLF, SSH, KTJ, DFK, JAT, PJH developed the design of the LIRE study. KTJ is the project director. All authors have read and approved the final version of the article. JGJ drafted the manuscript and all co-authors made contributions to revisions.

Acknowledgements

The study is supported by the National Institutes of Health 4UH3AR066795-02.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.cct.2015.10.003>.

References

- J.J. Jarvik, W. Hollingworth, P. Heagerty, D.R. Haynor, R.A. Deyo, The Longitudinal Assessment Of Imaging and Disability of the Back (LAIDBack) study: baseline data, *Spine* 26 (10) (2001) 1158–1166.
- M.C. Jensen, M.N. Brant-Zawadzki, N. Obuchowski, M.T. Modic, D. Malkasian, J.S. Ross, Magnetic resonance imaging of the lumbar spine in people without back pain [see comments], *N. Engl. J. Med.* 331 (2) (1994) 69–73.
- W. Brinjikji, P.H. Luetmer, B. Comstock, B.W. Bresnahan, L.E. Chen, R.A. Deyo, S. Halabi, J.A. Turner, A.L. Avins, K. James, et al., Systematic literature review of imaging features of spinal degeneration in asymptomatic populations, *AJNR Am. J. Neuroradiol.* (2014).
- J.G. Jarvik, L.S. Gold, B.A. Comstock, P.J. Heagerty, S.D. Rundell, J.A. Turner, A.L. Avins, Z. Bauer, B.W. Bresnahan, J.L. Friedly, et al., Association of early imaging for back pain with clinical outcomes in older adults, *JAMA* 313 (11) (2015) 1143–1153.
- J.M. Graves, D. Fulton-Kehoe, J.G. Jarvik, G.M. Franklin, Health care utilization and costs associated with adherence to clinical practice guidelines for early magnetic resonance imaging among workers with acute occupational low back pain, *Health Serv. Res.* 49 (2) (2014) 645–665.
- J.M. Graves, D. Fulton-Kehoe, J.G. Jarvik, G.M. Franklin, Early imaging for acute low back pain: one-year health and disability outcomes among Washington state workers, *Spine (Phila Pa 1976)* 37 (18) (2012) 1617–1627.
- M. Roland, M. van Tulder, Should radiologists change the way they report plain radiography of the spine? *Lancet* 352 (9123) (1998) 229–230.
- B.J. McCullough, G.R. Johnson, B.I. Martin, J.G. Jarvik, Lumbar MR imaging and reporting epidemiology: do epidemiologic data in reports affect clinical management? *Radiology* 262 (3) (2012) 941–946.
- J.G. Jarvik, R.A. Deyo, Diagnostic evaluation of low back pain with emphasis on imaging, *Ann. Intern. Med.* 137 (7) (2002) 586–597.
- J.G. Jarvik, W. Hollingworth, B. Martin, S.S. Emerson, D.T. Gray, S. Overman, D. Robinson, T. Staiger, F. Wessbecher, S.D. Sullivan, et al., Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial, *JAMA* 289 (21) (2003) 2810–2818.
- B. Martin, S.K. Mirza, J.D. Lurie, A.N.A. Tosteson, R.A. Deyo, Validation of an administrative coding algorithm to identify back-related degenerative diagnoses, International Society for the Study of the Lumbar Spine (ISSLS), May 14, 2013 (2013; Scottsdale, AZ).
- AMA - About CPT® <http://www.ama-assn.org/ama/pub/physician-resources/solutions-managing-your-practice/coding-billing-insurance/cpt/about-cpt.page>.
- J. Vogel, J.S. Brown, T. Land, R. Platt, M. Klompas, MDPHnet: secure, distributed sharing of electronic health record data for public health surveillance, evaluation, and planning, *Am. J. Public Health* 104 (12) (2014) 2265–2270.
- PopMedNet™ Distributed Research Network Technologies for Population Medicine <http://www.popmednet.org>.
- T.R. Ross, D. Ng, J.S. Brown, R. Pardee, M.C. Hornbrook, G. Hart, J.F. Steiner, The HMO research network virtual data warehouse: a public data model to support collaboration, *EGEMS (Wash DC)* 2 (1) (2014) 1049.
- J.G. Jarvik, B.A. Comstock, B.W. Bresnahan, S.S. Nedeljkovic, D.R. Nerenz, Z. Bauer, A.L. Avins, K. James, J.A. Turner, P. Heagerty, et al., Study protocol: the Back Pain Outcomes using Longitudinal Data (BOLD) registry, *BMC Musculoskelet. Disord.* 13 (1) (2012) 64.
- International Classification of Diseases, Ninth Revision (ICD-9) <http://www.cdc.gov/nchs/icd/icd9.htm>.
- B.I. Martin, M.M. Gerkovich, R.A. Deyo, K.J. Sherman, D.C. Cherkin, B.K. Lind, C.M. Goertz, W.E. Lafferty, The association of complementary and alternative medicine use and health care expenditures for back and neck problems, *Med. Care* 50 (12) (2012) 1029–1036.
- B.I. Martin, S.K. Mirza, G.M. Franklin, J.D. Lurie, T.A. MacKenzie, R.A. Deyo, Hospital and surgeon variation in complications and repeat surgery following incident lumbar fusion for common degenerative diagnoses, *Health Serv. Res.* 48 (1) (2013) 1–25.
- J.G. Jarvik, W. Hollingworth, P.J. Heagerty, D.R. Haynor, E.J. Boyko, R.A. Deyo, Three-year incidence of low back pain in an initially asymptomatic cohort: clinical and imaging risk factors, *Spine* 30 (13) (2005) 1541–1548 (discussion 1549).
- S. Bird, NLTK: The Natural Language Toolkit, "Proceedings of the COLING/ACL on Interactive presentation sessions, 2006.
- H. Harkema, J.N. Dowling, T. Thornblade, W.W. Chapman, ConText: an algorithm for determining negation, experienter, and temporal status from clinical reports, *J. Biomed. Inform.* 42 (5) (2009) 839–851.
- G. Gibson, Hints of hidden Heritability in GWAS, *Nat. Genet.* 42 (7) (2010) 558–560.
- NDC: <http://www.fda.gov/Drugs/InformationOnDrugs/UCM142438.htm>
- MED: <http://www.agencymeddirectors.wa.gov/opioiddosig.asp>.
- H. Quan, B. Li, C.M. Couris, K. Fushimi, P. Graham, P. Hider, J.M. Januel, V. Sundararajan, Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries, *Am. J. Epidemiol.* 173 (6) (2011) 676–682.
- J.N. Weinstein, T.D. Tosteson, J.D. Lurie, A.N. Tosteson, E. Blood, B. Hanscom, H. Herkowitz, F. Cammisia, T. Albert, S.D. Boden, et al., Surgical versus nonsurgical therapy for lumbar spinal stenosis, *N. Engl. J. Med.* 358 (8) (2008) 794–810.
- K. Thorpe, M. Zwarenstein, A. Oxman, S. Treweek, C. Furburg, D. Altman, S. Tunis, E. Bergel, I. Harvey, D. Magid, et al., A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers, *CMAJ* (2009).
- K. Loudon, S. Treweek, F. Sullivan, P. Donnan, K.E. Thorpe, M. Zwarenstein, The PRECIS-2 tool: designing trials that are fit for purpose, *BMJ* 350 (2015) h2147.
- PRECIS toolkit: <https://crs.dundee.ac.uk/precis/Help/Documentation/Toolkit>
- K.E. Johnson, C. Tachibana, G.D. Coronado, L.M. Dember, R.E. Glasgow, S.S. Huang, P.J. Martin, J. Richards, G. Rosenthal, E. Septimus, et al., A guide to research partnerships for pragmatic clinical trials, *BMJ* 349 (2014) g6826.
- W.W. Chapman, M. Fizman, B.E. Chapman, P.J. Haug, A comparison of classification algorithms to automatically identify chest X-ray reports that support pneumonia, *J. Biomed. Inform.* 34 (1) (2001) 4–14.
- P.C. Jacobs, W.P. Mali, D.E. Grobbee, Y. van der Graaf, Prevalence of incidental findings in computed tomographic screening of the chest: a systematic review, *J. Comput. Assist. Tomogr.* 32 (2) (2008) 214–221.