Research Aims

Aim 1: To determine whether inserting a description of age-specific prevalence of imaging findings among asymptomatic subjects into lumbar spine imaging reports decreases back-related interventions (imaging, injections, surgeries, etc.) over the subsequent year.

Aim 1a: To determine if inserting epidemiological evidence reduces Relative Value Units (RVUs) attributable to spine interventions (imaging, injections, specialist referrals, surgeries, etc.).

Aim 1b: To determine if inserting epidemiological data decreases opioid prescriptions.

Aim 1c: To determine if inserting epidemiological evidence decreases subsequent cross-sectional imaging (magnetic resonance (MR) and computed tomography (CT)).

Aim 1d: To explore whether adding epidemiological evidence decreases overall costs of care for low back pain based on CMS reimbursement.

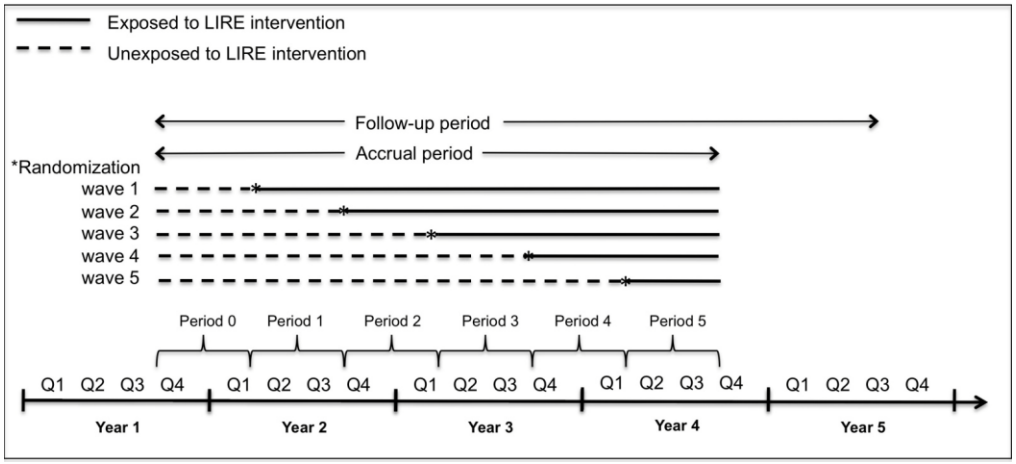
Aim 2: To determine whether inserting age-specific prevalence of imaging findings in asymptomatic subjects has a differential effect on subsequent back-related interventions if inserted into lumbar spine MR and CT imaging reports compared with plain films.

Aim 3: To determine if specific imaging findings influence subsequent interventions.

Intervention: This trial will study two groups of patients within providers, each will have had a lumbar imaging CPT code. Patients and providers at intervention clinics will receive additional prevalence summary data of incidental findings as a part of their radiology report. Control patients and providers will receive the standard imaging report without the LIRE text.

Design: Using a stepped wedge cluster randomized design[1], we will randomly assign all predetermined clinics at each site to receive the intervention at one of five fixed time-points. Interventions will roll out every six months at the start of the second quarter of UH3 Year 2 according to the schedule shown in Figure 1.

Figure 1: Proposed Randomization Schedule



During the UH2 project phase, we obtained a current and accurate enumeration of PCPs within clinics. Within each recruitment site, we sorted clinics by number of primary care providers into tertiles (e.g. small, medium, large clinics). From each tertile we will randomly select 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. Table 1

displays the site-specific strata definitions and size. In total, we will randomize 110 clinics with 1,824 PCPs as units of observation within those clinics. Note that we have chosen to use site-specific definitions for the size of the clinic with the goal of having balance of clinic size within each site. In addition, by balancing randomization on size we will be sure to have comparable time on control and intervention for each clinic size strata.

In the original project application, we assumed 128 clinics and 1,898 PCPs would participate in the LIRE project. After input from the Collaboratory Biostatistics Core, we excluded all clinics with a single PCP (n=18) from the primary study and statistical analysis and will only include clinics with 2 or more PCPs.

Table 1. Within-site stratified randomization schedule of clinics by number of PCPs.

Recruitment Site	Units of Randomization (# of PCPs)	PCP strata size boundaries (# clinics) Small	Medium	Large
Group Health Cooperative	25 (370)	5 to 10 (9)	11 to 15 (8)	16 to 41 (8)
Henry Ford Health System	26 (230)	3 to 6 (9)	7 to 9 (9)	10 to 24 (8)
Kaiser Permanent N. CA	20 (865)	17 to 29 (7)	33 to 39 (5)	43 to 106 (8)
Mayo Clinic Health System	39 (359)	2 to 4 (15)	5 to 9 (12)	11 to 34 (12)
Total	**110 (1,824)**			

Primary Outcome: We have devoted substantial effort towards developing and refining the primary outcome measure: a summary back-specific relative value unit (RVU). The back-specific RVU is a composite measure of spine intervention intensity that combines the overall intensity of resource utilization for back pain care into a single metric.

To develop the composite RVU measure, we used data from our large cohort of patients with back pain who comprise the Back pain Outcomes using Longitudinal Data (BOLD) Project, Agency for Healthcare Research and Quality (AHRQ)-funded study. During our work with the BOLD Project we developed algorithms to abstract electronic medical record (EMR) data across three health systems (two of which overlap with LIRE): Kaiser Northern California, Henry Ford Health System and Harvard Vanguard/Harvard Pilgrim. For the 5,239 BOLD cohort participants, we obtained extensive EMR data on pharmacy records, healthcare utilization (CPT codes), diagnoses and provider visits (ICD-9 codes), and inpatient hospitalization data.

Using the Medicare Physician Fee Schedule (http://www.cms.gov/) we generated and tested a mapping algorithm to assign more than 10,000 unique CPT codes to RVUs. A sample of RVUs from the 2012 CMS file is shown in Table 2. Using the BOLD cohort EMR data, we developed and tested an algorithm for aggregating individual RVUs across procedures over a time interval for a given patient, as well as across primary care providers or clinics.

Table 2. Example spine-related CPT codes and associated RVUs.

CPT Code	Description	RVU
72100	X-ray exam of lower spine - 2 views	1.07
97001	PT Evaluation	2.18
99214	Detailed office visit	2.26
99284	Emergency department visit - high moderate intensity	3.37
64483	Epidural injection for lumbar spinal stenosis	3.37
72131	CT lumbar spine w/o dye	6.27
72148	MRI Lumbar Spine w/o Contrast	11.31
63047	Removal of spinal lamina	32.89
22804	Fusion of the spine	71.60

To obtain a spine-related summary RVU from CPT and ICD-9 codes, we used an existing algorithm

developed by a colleague at Dartmouth College.[2-4] Aggregating across CPT codes identified by this algorithm yields the back-specific RVU.

We are currently preparing a manuscript describing this development work as well as a manuscript that directly influences and informs our LIRE UH3 proposal. Using BOLD cohort data, we identified a subset of patients who have had an early lumbar image (MRI/CT or plain film) following an office visit for back pain. Our BOLD cohort manuscript (in preparation) compares the one-year cumulative RVU of early-imaged patients to carefully matched BOLD cohort controls who did not have an early lumbar image. Preliminary results indicate a substantial downstream increase in healthcare utilization for patients who received an early image compared to propensity score matched controls. Patients who underwent a lumbar MRI or CT had a mean one-year RVU of 150 +/- 410, versus 120 +/- 450 for those who had an early plain film, versus 43 +/- 120 for carefully matched controls. Mapping the relative increases of utilization of nearly 80 and 110 RVUs for the plain film and advanced imaging modalities to the example codes shown in Table 2, we see that imaged patients undergo substantially more procedures. Our expectation for the UH3 project is that the insertion of normative prevalence data into lumbar imaging reports will reduce subsequent inappropriate healthcare utilization.

Secondary Outcomes: In addition to back-specific RVU, important secondary outcomes will be obtained and derived using electronic medical record data pulls and include: an indicator of opioid prescriptions written within 30 and 90 days after the index image (Aim 1b); subsequent cross-sectional re-imaging within 90 days and 12 months (Aim 1c); and medical costs (Aim 1d). In the BOLD project, we developed mapping algorithms based upon the United States Food and Drug Administration National Drug Codes (NDC)[5] that generate an indicator of whether or not an individual pharmacy record is an opioid analgesic. Similarly, we have enumerated and categorized a listing of CPT codes that indicate cross-sectional lumbar imaging (CT, or MRI).

General Analytic Strategy: To evaluate the effectiveness of inserting epidemiologic evidence into an imaging report we will use longitudinal regression methods such as linear mixed effects models (LMMs) or generalized linear mixed models (GLMMs) 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. However, correct model specification is required to ensure valid results when using LMMs or GLMMs and we will therefore use robust standard errors for our primary analysis. Therefore, we are effectively adopting a “working” correlation structure through the specification of flexible multilevel models (LMM or GLMM) but will rely on non-parametrically valid robust standard errors for inference where we cluster on the clinic. Secondary analysis will directly use generalized estimating equations (GEE) adopting simple exchangeable correlation models at the clinic level to determine whether conclusions appear sensitive to model specification.

In each analysis we will also consider a ‘washout period’ in the three months prior to the intervention being activated at a clinic, as determined by the randomization schedule. The rationale for a washout period is to reduce or eliminate within-provider cross-contamination of patient outcomes and utilization in the transition period between control and intervention. Including a washout period reduces the risk of having a patient initially treated in the control time period return to their primary care provider for subsequent care after the primary care provider has been exposed to the intervention through other patients. This reduces the potential bias due to within-provider cross-contamination of outcomes on the estimated intervention effect.

Primary Analysis: The primary longitudinal model for back pain specific RVUs will use a time-varying intervention status indicator $Status_{kt}$ (0 = control, 1 = intervention, for clinic k at time t). Use of the time-dependent intervention status indicator permits both within-clinic contrasts that inform intervention effects (post- versus pre-intervention) as well as contrasts across clinics with different intervention statuses within each time period. The specific regression model will adopt a functional form given below, with fixed effects for time (linear), age (18-39, 40-59, 60+, using two dummy variables), imaging modality type (plain film, CT, MRI using two dummy variables), and clinic size (small, medium, large, using two dummy variables), and site (Group Health Cooperative, Henry Ford, Kaiser Permanente, Mayo Clinic, using three dummy variables) in addition to random

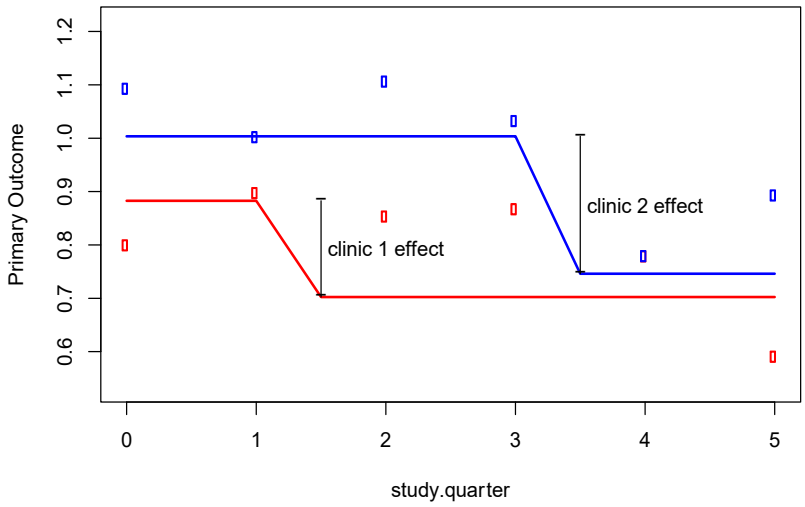
effects for provider, clinic, and intervention status:

$$Y_{ijk} = \beta_0 + \beta_1 \cdot Time_t + \beta_2^T \cdot Age_{ijk} + \beta_3^T \cdot Modality_{ijk} + \beta_4^T \cdot Size_k + \beta_5^T \cdot Site_k + \lambda_0 \cdot Status_{kt} + b_{k,0} + b_{k,1} \cdot Status_{kt} + a_{jk,0} + e_{ijk}$$

mean model
clinic random effects
provider random effects and errors

We will collect the outcome measure Y_{ijk} on patient i ($i= 1,2,\ldots,n_j$) under primary care provider j ($j =1,2,\ldots,n_k$) enrolled in time period t ($t = 0,1,2,\ldots ,5$) in order to evaluate the overall effect of the intervention at the level of the clinic k ($k = 1,2,\ldots,110$). Note that we will collect a single outcome measure for each subject recording the total utilization (RVU) over the 12 months after the index imaging event. Given that the random effects structure may contain additional elements (see below) we will use a robust standard error to test the null hypothesis that $\lambda_0 = 0$. For example, using SAS PROC MIXED we can use the "empirical" option in order to obtain robust standard errors. Alternatively, use of the jackknife (at the clinic level) provides a robust standard error estimate (if using R and *lmer*) that is simple to compute.

Figure 2: Hypothetical example of study data showing random clinic intercepts and intervention effects. Each line shows the expected profile for a specific clinic. Here clinic 1 initiates intervention at quarter=2 while clinic 2 initiates at quarter=4.



<u>*Key Model Parameters*</u>: The primary parameter of interest is λ_0, which represents the average effect of the intervention adjusting for temporal trends ($Time_t$), clinic characteristics ($Site_k$, $Size_k$), and individual covariates (Age_{ijk}, $Modality_{ijk}$). In order to interpret the random effects structure we focus on clinic level means removing covariate effects where we have: adjusted mean at clinic k for times prior to intervention = $\beta_0 + b_{k,0}$; and the adjusted mean at clinic k for times after start of the intervention = $\beta_0 + b_{k,0} + \lambda_0 + b_{k,1}$. For clinic-specific means we average over both providers ($a_{jk,0}$) and patients (e_{ijk}). Using this representation we interpret β_0 as the pre-intervention adjusted overall mean outcome averaging across all clinics, and $b_{k,0}$ is the difference between that adjusted overall mean and the pre-intervention (baseline) mean for clinic k. The variance, var($b_{k,0}$), is a measure of the variation in the baseline mean outcome across clinics. The change in the adjusted mean outcome for clinic k is given by: (post-intervention adjusted mean) – (pre-intervention adjusted mean) = $(\beta_0 + b_{k,0} + \lambda_0 + b_{k,1}) - (\beta_0 + b_{k,0}) = \lambda_0 + b_{k,1}$. Here λ_0 represents the average intervention effect across all clinics and $b_{k,1}$ represents the difference between that average intervention effect and the intervention effect for clinic k. The variance, var($b_{k,1}$), is a measure of the variation in the change associated with intervention across clinics, or a measure of the heterogeneity of the intervention effect.

Our primary regression model acknowledges the fundamental multilevel structure of individual-level data collected in health care systems with patients nested within providers, and providers nested within clinics. Although, the basic intervention contrast is the pre-post change associated with the initiation of intervention for each clinic, we do not propose using clinic-level summary measures for inference since the weighting of both patients and providers is not simple when heterogeneity of cluster sizes exists (e.g. PCPs per clinic, and patients per PCP). A proper multilevel model allows for optimal weighting based on the estimated variance components

(e.g. Gauss-Markov) and yields both an efficient summary of the overall intervention effect, as well an estimate of the variability in the magnitude of effect across clinics. However, we will not rely on the covariance model being correct for statistical inference and will use a robust (empirical) standard error. With greater than 100 total clusters (clinics) we expect valid inference and proper test size and do not anticipate needing to perform any correction such as the jackknife[7] (recommended when the number of clusters is small).

In our analysis we effectively assume that individual patients are nested within a single provider. However, in practice a patient may change providers during the follow-up year over which the primary outcome is captured. However, our basic mixed model covariance structure will simply use the assigned primary provider at the index imaging time. Therefore, we do not rely on model-based standard errors since the covariance structure may not match the true within-clinic covariance structure. We will use robust standard errors clustering at the clinic level, and therefore our analysis is valid even if there are changes in patient provider leading to an incorrectly specified covariance structure. Robust standard errors remain valid when a covariance model is not correctly specified. Furthermore, key secondary analysis of the primary outcome will directly use GEE and only cluster at the clinic level and provider level linkages are not used (nor needed) for simple GEE analysis.

Secondary Analyses of Primary Outcome: We will conduct additional secondary analyses that evaluate the sensitivity of the multilevel model to the assumed basic random effects structure. We have included in the primary model multilevel random intercepts and a random effect for the clinic-level intervention. However, we will expand the random effects structure to also permit random slopes on time for both clinics and providers. Given the relatively short duration of follow-up with only six (6) total measurement times we do not expect strong heterogeneity across providers or clinics in cluster-specific temporal trends. Figure 2 shows an example of hypothetical data series for two clinics (assuming aggregation of providers to a clinic summary) and illustrates both the staggering of the crossover time and the potential to observe clinic-specific intervention effects. This figure also illustrates the fact that separating random effects of time (linear) from random effects of intervention would be difficult since time and intervention status are correlated give the unidirectional crossover from control to intervention. In addition, we will use GEE as a covariance model robust inference method and therefore can produce valid point estimates and confidence intervals without relying on correct covariance specification. Details of model choice and comparison of alternative models for longitudinal cluster level crossover trials is presented in French and Heagerty (2008) and comparison of alternative approaches is recommended.

Models for time and intervention effect: Our primary analysis adopts a linear adjustment for calendar time in order to remove any large-scale temporal trends that may bias estimates of intervention effects. However, our basic regression model assumes a common (adjusted) mean for all times after the initiation of intervention. In practice there may be a delay in the impact of intervention so alternative models will be considered that incorporate a delayed and/or gradual effect of intervention. For example, the basic coding of the time-dependent variable $Status_{kt}$ takes the value 0 pre-intervention and the value 1 post-intervention. Delay in the impact of intervention can be accommodated using alternatives such as: 0 pre-intervention; 0.5 for quarter 1 after intervention; and 1 for all other post-intervention quarters. Such a modified model would allow full impact of the intervention to require two quarters of exposure. We will conduct secondary analyses to explore alternative models for the accumulation or delay of the intervention effect.

Secondary Outcome Analysis: We will also analyze the impact of intervention on the rate of opioid prescription using Generalized Linear Mixed Models (GLMMs). For Aim 1b let Y_{ijk} =1 if opioids were prescribed within a given timeframe (e.g. 30 days or 90 days) to patient i (i= 1,2,…,n_j) seen by primary care provider j (j =1,2,…,n_k) within clinic k (k = 1,2,…,110). Analysis for this outcome will use a logistic mixed model given as:

$$logit(p_{ijk}) = \beta_0 + \beta_1 \cdot Time_t + \beta_2^T \cdot Age_{ijk} + \beta_3^T \cdot Modality_{ijk} + \beta_4^T \cdot Size_k + \beta_5^T \cdot Site_k + \lambda_0 \cdot Status_{kt} + \quad \text{mean model}$$
$$b_{k,0} + b_{k,1} \cdot Status_{kt} + \quad \text{clinic random effects}$$
$$a_{jk,0} \quad \text{provider random effects}$$

where p_{ijk} denotes the probability that Y_{ijk} =1. Our secondary outcome analysis parallels the primary and will be based on a natural multilevel mixed model, with additional robust secondary analysis provided by GEE. For Aim 1c we will use Y_{ijk} =1 if CT or MR imaging occurs within a specified timeframe (e.g. 90 days or 12 months) after the index imaging event.

Medical costs (Aim 1d): Spine-related costs of care will be estimated using two approaches. First, we will use the spine-related RVU calculated in Aim 1a and estimate clinical-level, spine-intervention expenditures using the annual Medicare-determined payment amount per RVU (e.g., CY2013 = $34.023 per RVU). (reference: http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-LN/MLNProducts/downloads/medcrephysfeeschedfctsht.pdf)

Second, as a proxy for costs of spine care, we will use a standard set of reimbursement amounts, i.e., CMS-based payments, and estimate clinic-level spine-related aggregate expenditures by applying CPT-based payment amounts to specific spine-intervention events (e.g., imaging, office visits, procedures, other). We will present monthly and annual means, medians, and ranges of clinic-level cost estimates, prior to and subsequent to implementing the epidemiological intervention. We will assess the level of right-skewness in the expenditure estimates and use t-tests to compare arithmetic means for clinic-level expenditures. In the case of considerable skewness, we will test for differences in logarithmically transformed mean clinic-level expenditures (before and after implementing intervention). We will also describe categories of prescriptions ordered, when available in the electronic medical records for a health system, and estimate costs for prescribed spine-related medications

Analysis for Aim 2: The hypothesis of Aim 2 is that there will be a differential effect of the intervention according to the imaging modality used. 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 two indicator variables coding CT and MR, with plain film as the reference) and $Status_{kt}$. A test of the interaction terms (2 degree of freedom Wald test) will be used to test the null hypothesis that the effect of the intervention does not vary according the imaging modality.

Analysis for Aim 3: The hypothesis of Aim 3 is that there will be a differential effect of the intervention according to the results that are found in the imaging report. We will use an additional variable, $ImageFinding_{ijk}$, that takes the value 1 if a significant image finding is present, and 0 otherwise (see detail regarding variable specification in protocol). We will test the null hypothesis that the interaction between $ImageFinding_{ijk}$ and $Status_{kt}$ is zero using a Wald test.

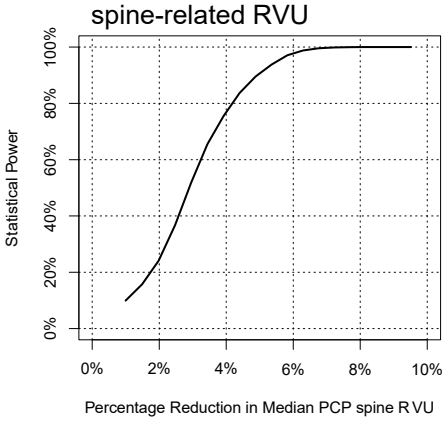
Power Calculations: Our UH2 efforts with respect to sample size and statistical power focused on two key items. First, an important aim of our UH2 Working Group 2 was to obtain an accurate clinic and provider count for each health system. We will now randomize n=110 clinics (1,824 PCPs), which is slightly lower than the n=128 clinics (1,898 PCPs) assumed in our initial project application. However, the majority of the clinics that were dropped were those with only one PCP and therefore would not have contributed much information to the analysis. Second, in our Working Group 3 we sought to develop and characterize a composite RVU summary to be used as the primary outcome measure in this study. In our UH2 project application we discussed statistical power in the context of an important secondary outcome measure, a reduction in subsequent opioid prescription rates. We now present statistical power for the primary outcome measure using data from the BOLD Registry to inform key design parameter estimates.

To our knowledge, off-the-shelf calculators do not exist that would adequately characterize statistical power for a stepped wedge cluster randomized trial with a varying number of sampling units between clusters. We therefore utilized simulation methods to generate and analyze data that closely mimics the design characteristics we anticipate for this study. With a simulation approach, we were able to include estimates of both patient and clinic-level variability and implement the proposed primary analysis methods: random intercept linear mixed effects models for RVU outcomes; and generalized linear mixed models for opioid prescription rates. All simulations

were conducted using R (version 3.0.1) with the *lmer* and *glmmPQL* functions implementing mixed model estimation.

Power for Primary Outcome: In the BOLD Registry we identified 639 patients in the Kaiser Permanente and Henry Ford health systems that had a qualifying lumbar image within 6 weeks of a PCP visit, the majority (74%) of which occurred within 7 days. As one would expect with a measure of health care utilization intensity, patient-level RVUs are positively skewed. In our simulations and in the future analysis of study data, we therefore utilize an approximately normalizing transformation of log(RVU + 1) but will make interpretations regarding effect size back on the original RVU scale.

Figure 2: Statistical power for spine-related RVU



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.

The numbers of clinics and primary care providers were considered fixed 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 1,000 simulated data sets and performed mixed model estimation with each data set. In Figure 2, we show statistical power for the primary outcome measure of PCP spine-related RVU under the proposed study design. The study has greater than 90% power to detect reductions in the median spine-related RVU of 5.0% or larger. For a patient receiving a lumbar CT, a 5% reduction in spine RVU translates into one fewer additional lumbar CT scan on average compared to a patient unexposed to the LIRE intervention.

Power for Secondary Outcome: Using the updated clinic and provider listing, we repeated the UH2 power analyses for a reduction in subsequent opioid prescriptions. We again used an average baseline opioid prescription rate of 22%, suggested from the pilot manuscript[9] to anchor the effect size of percent reduction in the baseline rate. A clinic-specific baseline opioid prescription rate was drawn from a Beta(α=6, β=20) distribution; we used this rate to draw a baseline opioid prescription rate random effect for each primary care provider using a clinic-specific log-normal distribution. We generated 1,000 simulated data sets using the effect size of a 7.5% reduction in the opioid reported in our UH2 application and evaluated each using a GLMM assuming random intercepts. With 110 clinics randomized, the study remains well powered (88.9% power) to detect a reduction in the rate of opioid prescriptions of 7.5% or larger (e.g. 22% down to 20.4%).

Statistical Analysis Plan References

1. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007;28:182-91.
2. Martin B, Mirza SK, Lurie JD, et al. Validation of an administrative coding algorithm to identify back-related degenerative diagnoses. *International Society for the Study of the Lumbar Spine (ISSLS)*. Scottsdale, AZ, 2013.
3. Martin BI, Gerkovich MM, Deyo RA, et al. The association of complementary and alternative medicine use and health care expenditures for back and neck problems. *Med Care* 2012;50:1029-36.
4. Martin BI, Mirza SK, Franklin GM, et al. Hospital and surgeon variation in complications and repeat surgery following incident lumbar fusion for common degenerative diagnoses. *Health Serv Res* 2013;48:1-25.
5. http://www.fda.gov/drugs/informationondrugs/ucm142438.htm, accessed September 13, 2013.

6. Diggle PJ, Heagerty PJ, Liang KY, Zeger SL. Analysis of Longitudinal Data. Second Edition ed: Oxford University Press; 2002.
7. Efron B, Stein C. The jackknife estimate of variance. *Annals of Statistics*. 1981;9(3):586-96.
8. French B, Heagerty PJ. Analysis of longitudinal data to evaluate a policy change. *Statistics in Medicine* 2008; 27(24):5005-25.
9. McCullough BJ, Johnson GR, Martin BI, Jarvik JG. Lumbar MR imaging and reporting epidemiology: do epidemiologic data in reports affect clinical management? Radiology. 2012;262(3):941-6. PMCID: 3285226

Lumbar Imaging with Reporting of Epidemiology (LIRE) Statistical Analysis Plan Revisions

The following table includes amendments to our statistical analysis plan (SAP) that we included in statistical models for the manuscript but that we did not specify in our original SAP.

Amendments to Statistical Analysis Plan (SAP)	
Number of clinics	
• 110 to 98 clinics	When we wrote the SAP we estimated that we would have 110 clinics participating in the trial. We dropped ten clinics prior to randomization for logistical reasons. We dropped two additional small clinics just after randomization and before the first data submission: one had closed and the other had been subsumed in another clinic.
Covariates	We made all decisions regarding covariates before we began the statistical analysis.
Covariates added	*Rationale*
• Site by time interactions	Allow for differential time trends by site
• Gender	Allow for outcome differences by gender (precision)
• Charlson Comorbidity Index category	Comorbidity measures are associated with outcomes. We sought greater precision by including this general measure of health.
• Prior opioid status (opioid models only)	Allow for differences in outcome between those without and those with prior opioid prescriptions (precision)
Covariates transformed	
• Time	The SAP and protocol paper indicated that we would model time linearly as 6-month time period (t = 0, 1, 2, 3, 4, 5). Upon further thought, we decided that a more granular measure (i.e. day) would be more appropriate.
• Clinical importance of imaging findings	To assess the impact of image findings on the treatment effect, the SAP and protocol paper specified that a binary measure indicating the presence of a clinically important finding would be used. We instead decided to use a three-category measure so that we could measure effect modification among the following groups: patients with likely clinically important findings, patients without likely clinically important findings but with one of the findings included in the LIRE intervention text, and patients without either of these types of finding.

Outcomes	
• Opioid prescriptions	The original SAP defined the opioid outcome as a binary indicator of opioid prescriptions within 90 days, which is the same outcome that we use in the manuscript. When we were writing the protocol paper we thought that morphine equivalent dose (MED) would be more sensitive to the intervention since it could capture reductions in the amount of opioids within prescriptions. However, we discovered that some sites had high missingness for prescription characteristics that we needed to calculate MED. We also encountered challenges identifying duplicate prescriptions. We concluded that it would be better to revert to our original opioid measure for which we would have considerably greater accuracy.