Estimating the effect of healthcare interventions on the distribution of health state severity for low back pain in the Global Burden of Disease study

Caroline Purcell

A thesis

Submitted in partial fulfillment of the

Requirements for the degree of

Master of Public Health

University of Washington

2019

Committee:

Theo Vos

Stein Emil Vollset

Program Authorized to Offer Degree:

Department of Global Health

University of Washington

Abstract

Estimating the effect of healthcare interventions on the distribution of health state severity for low back pain in the Global Burden of Disease study

Caroline Purcell

Chair of the Supervisory Committee:

Theo Vos

Department of Global Health

Background: Many causes of non-fatal health burden present with varying degrees of severity from asymptomatic to most severe. The severity distribution for diseases amenable to healthcare interventions is expected to vary as function of healthcare access and quality (HAQ). Current methods used in the Global Burden of Disease (GBD) study assume a constant severity distribution over space and time, ignoring any effect of healthcare interventions. This paper presents a method to incorporate information on the effect of healthcare interventions on health state severity and quantify the relationship between health state severity and HAQ in order to generate location-specific estimates of average condition severity. Using low back pain (LBP) as an example condition, estimates of intervention efficacy and utilization are generated and used to estimate averted and avoidable burden.

Methods: Healthcare interventions for LBP were identified from the Cochrane Database of Systematic Reviews. Efficacy was assessed in terms of the standardized mean difference in disability relative to some usual care or placebo group. Interventions were grouped into five intervention classes: (1) surgical; (2) behavioral, cognitive, and physical therapies; and three classes of analgesics ((3) NSAID, (4) opioid, and (5) non-opioid non-NSAID analgesics). Effect sizes were pooled across all interventions in a class using a network meta-analysis framework. The overall treatment effect for LBP was calculated as the utilization-weighted sum of the intervention class effect sizes. The effect of treatment for LBP was also calculated assuming an aspirational 100% utilization of the optimal set of interventions among all

individuals with LBP. The overall treatment effect was applied to the GBD LPB disability weight distribution generated from the United States based Medical Expenditure Panel Survey in order to estimate the relationship between intervention utilization and average disability. Using the Healthcare Access and Quality Index (HAQI) as a proxy for access to interventions for LBP, the relationship between HAQI and average LBP disability per case was linearly interpolated. For each country in the GBD, current YLDs, averted YLDs, avoidable YLDs, and YLDs for the optimal treatment scenario were calculated.

Results: A total of 134 trials representing 160 unique intervention-comparison combinations were analyzed. Surgical interventions and NSAIDs were the most effective interventions (SMDs -0.44 (-0.70, -0.20) and -0.28 (-0.52, -0.04) respectively). The overall effect of healthcare interventions on LBP disability was estimated to be -0.17 (-0.35, 0.01) for LBP without leg involvement and -0.29 (-0.56, -0.01) for LBP with leg involvement. The maximum achievable proportion of LBP burden avoided through use of healthcare interventions under routine health care circumstances was 24.4% (1.4-41.3%) based on the relationship between HAQI and LBP disability. A hypothetical 100% utilization of the optimal treatment could avoid an additional 22.6% (4.1-31.2%) of LBP burden leaving 53.0% (28.0-95.4%) of LBP burden that cannot be addressed using existing healthcare interventions for LBP.

Interpretation: Estimation of the relationship between intervention efficacy, utilization, and LBP severity indicates that health interventions impact LBP severity and imply that LBP severity should vary according to health system quality and intervention utilization. The methods presented in this study represent a generalizable approach to estimate location-specific severity distributions. Applied across the GBD, this method would allow for the estimation of averted and avoidable burden which serve as useful benchmarks for funders looking to expand coverage of and access to healthcare interventions. A reduction of the remaining burden which cannot be addressed using current healthcare technologies would require research and development to create novel approaches to treatment.

Introduction

Background

The Global Burden of Disease (GBD) study is the only comprehensive, internally consistent study of the world's health to date. Through an iterative estimation process, the study incorporates information on mortality and morbidity to systematically estimate the burden of diseases, risk factors, and causes of death according to age, year, sex, and country. The GBD qualifies health loss in terms years lived with disability (YLDs), which capture time lived with reduced functional health status, and years of life lost (YLLs), which measure health loss resulting from premature mortality. YLDs and YLLs are summed into disability-adjusted life years (DALYs), which serve as a uniform metric for the comparison of burden by cause across time and severity.[1]

Each cause of nonfatal disease burden within the GBD framework is associated with one or more consequence(s) of a disease or injury, termed sequela(e). Taken together, the sequelae included in the GBD study represent a mutually exclusive and collectively exhaustive list of all known causes of disease and disability. For each disease, the associated sequelae are usually, but not always, consistent with clinical or physiological classifications of disease severity. Each sequela captured in the GBD is associated with a health state description and disability weight (DW). Health state descriptions provide simple explanations of the symptoms and functional outcomes of each health state and were used as the basis for generating DW values. DWs serve as the basis for estimation of YLDs and are used to weight the contribution of all sequelae to the overall burden attributable to a given disease. DWs are formulated on a scale from 0-1, with 0 representing no functional health loss, and 1 being equivalent to death. YLDs are thus generated by multiplying the number of individuals experiencing a given sequela by its associated DW.[2]

Purpose

For some causes in the GBD sufficient epidemiological data is available to model each related sequela independently; however, many others lack detailed information on the occurrence of individual sequela either because these quantities are not frequently reported in the literature or because studies report the occurrence of clinical classifications of disease that do not correspond to the categorizations of functional health status used in the GBD study. In such cases, GBD estimates of the sequela-specific burden are generally derived by modeling the proportion of total disease prevalence and incidence attributable to each sequela, referred to as the severity distribution, and multiplying the overall disease occurrence by these severity distribution proportions.[3] Currently, the information used to estimate the distribution of health state severity is quite restricted in terms of geographic coverage for many causes within the GBD. As a result of this limited scope of information, severity distributions are location-invariant for many of the large causes of disability within the GBD. That leads to estimates of YLDs for these causes that are implausibly constant across geographies even if they are highly amenable to care and would thus be expected to vary as healthcare access and quality improves.

The purpose of this project is to develop a method to quantify the effect of healthcare interventions on health state severity using low back pain (LBP) – the largest contributor of YLDs in the GBD – as an example condition. This approach incorporates information from randomized controlled trials evaluating the effect of existing healthcare interventions on condition-specific disability, which have previously not been considered within the context of the GBD. It also allows for the estimation of the relationship between health state severity and healthcare access and quality, which can be used to generate location-specific estimates of avoided and avertable burden under current and optimal access scenarios. The work provides a useful benchmark for funders looking to expand coverage of and access to healthcare interventions and highlights the proportion of burden which is unavertable using existing healthcare technologies thus underscoring the role of research and development in creating new and more efficient modes of treatment. By breaking down non-fatal burden into that which has been averted, that which could be avoided, and that which cannot be impacted upon using existing technologies, this work provides a roadmap for clinicians, public health practitioners, and policy makers alike to translate GBD findings into actionable results.

Methods

Overview

Healthcare interventions for LBP were identified and information on the effect of each intervention on LBP severity was extracted from Cochrane Library systematic reviews. Interventions were grouped into classes and effect sizes were pooled across all interventions in each class using network meta-regression. A combined effect size was calculated as the utilization-weighted sum of the intervention class effect sizes. An alternative combined effect size was also calculated for a scenario assuming full utilization of the optimal treatment (FUOT). The overall effect size was applied to the GBD severity estimates to estimate the average DW of LBP in the absence of health care interventions (intervention-deleted severity). The average DW for the FUOT scenario was estimated by applying the FUOT intervention effect to the intervention-deleted severity estimates. The Healthcare Access and Quality Index (HAQI)[4] was used as a proxy for intervention utilization, and the relationship between HAQI and average disability was estimated, assuming that HAQI in the zero-utilization scenario is 0. For each country, current YLDs, averted YLDs, avoidable YLDs, and YLDs for the FUOT scenario were calculated based on GBD LBP prevalence estimates and the average DW per case for each scenario predicted based on HAQI. A high-level overview of study methods is presented in Figure 1.

Severity distribution inputs

Data sources and methods used to estimate the health state severity distributions for LBP in the GBD have been described elsewhere.[3,5] Briefly, information on heath state severity for LBP was obtained from the US Medical Expenditure Panel Survey (MEPS).[6] MEPS is an annual survey conducted among the non-institutionalized adult, US population constructed of overlapping panels, each of which is composed of five rounds of data collection and spans two years. MEPS collects information on respondents' health system interactions, including self-reported health conditions, health service utilization, and healthcare expenditures. During rounds two and four of each panel, respondents are asked to complete the 12-item Short Form Health Survey (SF-12), which measures functional health status over the past four weeks. MEPS data from 2000-2013 was included in this analysis, covering MEPS panels 4-17 (SF-12 administration did not begin until 2000 so panels 1-3 were excluded from analysis).

To map composite SF-12 scores to GBD DWs, 62 health state descriptions were selected which represented the entire spectrum of health state DWs from least to most severe. A convenience sample of respondents was asked to fill out an SF-12 form for a hypothetical individual experiencing the described health state. A loess regression was fit on the survey data to develop a mapping from composite SF-12 score to GBD DW. This function was used to predict the DW for each MEPS respondent on the basis of their composite SF-12 score. Logit-transformed predicted DWs were regressed on health conditions for each individual using a mixed-effects model with fixed effects on each health condition and a random effect on respondent. This regression was used to derive comorbidity-corrected health state DWs for individuals with LBP with leg pain (ICD-9 code 724) and LBP without leg pain (ICD-9 code 722). Comorbidity-corrected DW predictions were generated 1,000 times, drawing from the variance-covariance matrix.[3,5] For each draw, the average DW was calculated across the sample of individuals with LBP. To arrive at a final summary estimate of the average DW per LBP case, the set of 1,000 draws was collapsed to the mean and standard deviation.

Treatment effect estimation

Information on the efficacy of health care interventions for LBP was obtained from the Cochrane Database of Systematic Reviews. Reviews catalogued in the Cochrane database have been conducted using a common, systematic approach meant to comprehensively evaluate available empirical evidence on topics related to health care.[7] A title/abstract keyword search was conducted for LBP ("back pain" OR "lumbar pain" OR "back ache" or "lumbago). All returned records were title screened for inclusion. Reviews that assessed interventions in subpopulations (e.g. pregnant women, surgery patients), those that assessed treatments for conditions other than LBP, and those that did not explicitly review health care interventions or treatments (e.g. review of diagnostic criteria, prescription adherence, professional practice guidelines) were excluded. All remaining reviews were full-text screened. Reviews that only

presented qualitative analysis of treatment effects, did not include outcomes relevant to LBP severity, and/or did not evaluate any outcomes on a continuous scale were excluded. For each study included in the review, information on intervention type, intervention description, comparison type, comparison description, outcome scale/instrument, follow-up duration, and intervention and comparison group sample size, outcome mean, and outcome standard deviation was extracted for each outcome related to disease severity. Information on other outcomes not related to severity such as adverse events, return to work, and physician assessment of recovery was not extracted. Interventions were grouped into six classes based on clinical relevance: surgery; behavioral, cognitive, and physical therapies; NSAID pharmaceuticals; opioid analgesics; non-opioid non-NSAID analgesics; and a reference category of usual medical care or placebo treatment. If multiple follow-up durations were reported for a single trial, the longest available duration was extracted. The optimal treatment for LBP was designated as physical and psychological therapies and NSAID use per the clinical practice guidelines published by the American College of Physicians.[8] Surgery was also included in the optimal treatment scenario for the small subset of cases of LBP with leg pain resulting from disc problems or spinal stenosis which do not resolve over time or following treatment with other non-surgical interventions .[9]

The main outcomes assessed by reviews were disability and pain, with few studies administering the SF-12. Disability was chosen as the outcome for analysis as it is more comparable than pain to a quality of life measure like SF-12 score. Efficacy was defined in terms of the standardized mean difference (SMD) in disability outcome measure between the intervention and comparison groups. For each intervention-comparison pair, the SMD was calculated based on sample size, outcome mean, and outcome standard error using the metafor package in R version 3.6.0.[10,11] Effect sizes were pooled using a newly-developed Bayesian mixed-effects meta-regression tool. The tool estimates between-study heterogeneity and adds this variation back to the estimated effect and can incorporate Gaussian priors for any estimated parameter. The model uses a trimmed maximum likelihood estimator which excludes outlier observations based on their contributions to the likelihood function. The degree of trimming is moderated by a user-specified percentage of observations to be trimmed. Uninformative priors were set on all parameters and the trimming percentage was set to 10%. Data were analyzed as a network in order to synthesize all available information across direct and indirect comparisons of each intervention class compared to the reference group (general care/placebo).[12]

Utilization estimates

Utilization estimates were extracted from the MEPS household component event files which contain information related to medical events for each respondent. Event files are available for prescription medicines, dental visits, office-based medical provider visits, hospital inpatient stays, emergency room visits, outpatient department visits, home health events, other medical expenses. Each person can report multiple medical conditions each of which can be associated with multiple medical events. To estimate utilization of healthcare interventions for LBP, only those events associated with LBP were assessed with the exception of pharmaceutical events which were assessed among all individuals with LBP regardless of the condition(s) associated with the event.

Utilization of each pharmaceutical intervention class (NSAIDs, opioids, and non-opioid non-NSAID analgesics) was assessed for each individual from the MEPS prescription medicines file. This file reports each prescription medication obtained by a respondent over the panel which are classified into up to three therapeutic classes according to Multum Lexicon labels from the Cerner Mulnum, Inc. drug database. Therapeutic class variables were used to identify prescriptions for NSAIDs, opioids, and non-opioid non-NSAID analgesics. Individual utilization for each pharmaceutical intervention class was defined as use of least one medication in that intervention class at any point during the round.

Utilization of surgical interventions for LBP was assessed using the MEPS office-based medical provider visits, emergency room visits, outpatient department visits, and hospital inpatient stays files. The office-based, outpatient, and emergency files contain variables indicating if an individual received anesthesia and if a surgical procedure was performed during the visit. The hospital inpatient stays file has a variable

indicating whether any operation or procedure was performed while the person was in the hospital. Individual utilization of surgical interventions was defined as a least one outpatient visit, office-based visit, or emergency event coded to LBP during which the individual received anesthesia and a surgical procedure and/or at least one hospital inpatient stay coded to LBP during which the individual received any operation or procedure.

Utilization of behavioral, cognitive, and physical therapies was assessed using the MEPS office-based medical provider visits, outpatient department visits, and home health files. The office-based and outpatient visit files contain variables indicating if an individual received physical therapy, occupational therapy, or psychotherapy during their visit. The home health file contains variables indicating if an individual received physical or occupation therapy. Individual utilization of behavioral, cognitive, and physical therapies was defined as at least one office-based or outpatient visit coded to LBP during which the individual received physical therapy, occupational therapy, or physical therapy and/or at least one home health visit during which the individual received physical or occupational therapy.

Overall utilization of each intervention class was calculated as the proportion of individuals with LBP utilizing that intervention out of all individual with LBP across all panels. As severity distribution inputs were estimated separately for individuals with LBP with and without leg pain, utilization was also tabulated separately for these two conditions.

Effect sizes for each intervention class were multiplied by the corresponding utilization estimates and summed to calculate and an overall intervention effect size for LBP (equation 1)

$$SMD_{overall} = \sum_{i=1}^{n} SMD_i * coverage_i \quad (1)$$

where n represents the number of intervention classes. Separate intervention effects were calculated for LBP with and without leg pain. This approach requires the assumption the effect of each intervention is independent of the effect of any other class (i.e. no interaction between intervention classes). The standard deviation of the overall effect sizes was calculated as the square root of the summed variances of each individual intervention effect as follows

$$\sigma_{overall} = \sqrt{\sum_{i=1}^{n} \sigma_i^2} \quad (2)$$

The FUOT effect size was calculated in the same way, assuming 100% utilization of the optimal set of interventions for LBP and setting utilization of all other interventions to 0%.

Burden estimation

The average DW under the intervention-deleted scenario was calculated by applying the overall intervention effect to the comorbidity-corrected DWs calculated from the MEPS sample of individuals with LBP with and without leg pain. Comorbidity-corrected DWs were translated back into their corresponding SF-12 score based on the loess regression fit on composite SF-12 and GBD DW described above. The standard deviation of these SF-12 scores was calculated, multiplied by the overall intervention effect, and added to each individual SF-12 score to calculate the predicted SF-12 score under the intervention deleted scenario according to equation 3.

$$SF12_{intervention-deleted,i} = SF12_i + SMD_{overall} * \sigma_{SF12} \quad (3)$$

For the FUOT scenario, the standard deviation of SF-12 was multiplied by the FUOT intervention effect and subtracted from the intervention-deleted SF-12 values to calculate the predicted SF-12 score under full utilization of the optimal treatment for LBP.

$$SF12_{FUOT,i} = SF12_{intervention-delted,i} - SMD_{FUOT} * \sigma_{SF12} \quad (4)$$

Intervention-deleted and FUOT SF-12 scores were then translated back into DWs and average across the entire sample.

The average DW calculated for the MEPs sample was assigned the HAQI value for the United States in 2007, the midyear of the MEPS input data, while the HAQI value for the intervention-deleted scenario was assumed to be 0. The association of average DW and HAQI was assessed using simple linear interpolation. The average DW per LBP case was calculated based on the 2017 HAQI values for all 195 countries included in the GBD study.

Averted burden was defined as YLDs that have been prevented through existing health care intervention efforts. Avoidable burden was defined as YLDs that theoretically could be prevented through improvements in intervention utilization and quality. Using GBD estimates of LBP prevalence for each country, *c*, in 2017, current YLDs ($YLDs_{current}$) were calculated as prevalence multiplied by the predicted average DW ($\widehat{DW}$) (equation 5).

$$YLDs_{current,c} = LBP\ prev_c * \widehat{DW_c} \quad (5)$$

Averted burden was calculated as the difference between expected YLDs using the average DW predicted at an HAQI of 0 ($\widehat{DW}_{HAQI=0}$) and current YLDs (equation 6).

$$YLDs_{averted,c} = LBP\ prev_c * \widehat{DW}_{HAQI=0} - YLDs_{current,c} \quad (6)$$

Avoidable burden ($YLDs_{avoidable}$) was calculated as the difference between current YLDs and expected YLDs using the average DW predicted for the highest observed HAQI value among all countries in 2017 ($\widehat{DW}_{highest\ HAQI}$) (equation 7).

$$YLDs_{avoidable,c} = YLDs_{current,c} - LBP\ prev_c * \widehat{DW}_{highest\ HAQI} \quad (7)$$

Avoidable burden under the FUOT scenario ($YLDs_{avoidable,FUOT}$) was calculated as the difference between expected YLDs using the average DW predicted ($\widehat{DW}_{highest\ HAQI}$) and expected YLDs using the average DW predicted for the FUOT scenario ($\widehat{DW}_{FUOT}$) (equation 8).

$$YLDs_{avoidable,FUOT,c} = LBP\ prev_c\ (\widehat{DW}_{highest\ HAQI} - \widehat{DW}_{FUOT}) \quad (8)$$

The remaining burden which cannot be intervened upon using existing healthcare technologies ($YLDs_{remaining}$) was calculated as the difference between current burden and burden under the FUOT scenario (equation 9).

$$YLDs_{remaining,c} = YLDs_{current,c} - YLDs_{avoidable,FUOT,c} \quad (8)$$

Results

Intervention effect

A total of 107 Cochrane reviews were returned from an initial search of the Cochrane Systematic Reviews Database, 67 of which met the criteria for full-text review. Of the remaining records, 31 met full-text inclusion criteria representing a total of 252 unique trials. Reviews comparing treatments in the same intervention class (e.g. one surgical procedure vs. another) were excluded, leaving a total of 134 trials for analysis (Figure 2). The 134 trials represented a total of 160 unique intervention-comparison combinations, the majority of which were comparisons of behavioral, cognitive, and physical therapies to usual care. Table 1 records the number of trials assessing each intervention-comparison pairing included in the network meta-analysis. Follow-up durations varied greatly, from immediately following treatment to 2 years post-intervention. The Roland Morris Disability Questionnaire and the Oswestry Low Back Disability Questionnaire were the two most common instruments used to assess disability and were reported in 69 and 46 of the 160 trial, respectively.

Figure 3 - Figure 10 show the mean and SD standardized mean differences for each study contributing to each comparison as well as pooled estimates of the standardized mean difference predicted from the

network meta-analysis. A negative standardized mean difference indicates that the intervention group reported less disability related to their LBP than the comparison group on average. The only study comparing behavioral, cognitive, and physical therapies to non-opioid, non-NSAID analgesics was trimmed (Figure 6) as the effect size reported from this study was in the opposite direction of that which would be expected: comparisons of behavioral, cognitive, and physical therapies to the reference category (Figure 3) indicate that these interventions are more effective than usual care while comparisons of non-opioid, non-NSAID analgesics to the reference category show no effect (Figure 8). Based on the properties of network meta-analysis, it should be expected that behavioral, cognitive, and physical therapies should be more effective than non-opioid non-NSAID analgesics.

Standardized mean differences of 0.2 are generally considered small, 0.5 medium, and 0.8 large.[13] Table 2 shows the estimated standardized mean difference and the corresponding confidence interval both with and without added uncertainty for unexplained between study heterogeneity for each intervention class compared to the usual care or placebo reference group. Non-opioid non-NSAID analgesics were the only intervention class that demonstrated no significant effect compared to placebo before adding additional uncertainty from estimated between study heterogeneity (0.02 [-0.16, 0.20]). The effects of opioid analgesics and behavioral, cognitive, and physical therapies were medium but not significant when accounting for inter-study variance (-0.21 [-0.43, 0.01] and -0.22 [-0.44, 0.01], respectively). NSAIDs had a small but significant effect, with an estimated standardized mean difference of -0.28 (-0.52, -0.04) compared to placebo, and surgical interventions were the most effective intervention class with an estimated effect size of -0.44 (-0.70, -0.20).

Intervention utilization

There were 17,237 MEPS respondents with LBP without leg involvement and 4,061 with leg involvement across 14 panels included in the analysis. Utilization of all interventions was higher among individuals with LBP with leg involvement than LBP without leg involvement. Figure 11 shows intervention utilization across each MEPS panel. NSAIDs and behavioral, cognitive, and physical therapies were the most utilized interventions in both groups. Surgical interventions were not highly utilized by individuals with LBP without leg involvement while non-opioid, non-NSAID analgesics were least utilized by those with LBP with leg involvement. These results were collapsed into a single estimate of utilization for each intervention to match the MEPS severity distribution inputs which were also pooled across all panels. Estimated utilization of behavioral, cognitive, and physical therapies, surgical interventions, opioid analgesics, NSAIDS, and non-opioid, non-NSAID analgesics was 20.7% (20.1 – 21.3%), 2% (1.8 – 2.2%), 12.4% (11.9 – 12.9%), 34% (33.3 – 34.7%), and 6.9% (6.5% – 7.3%) for respondents with LBP without leg involvement and 31.1% (29.7 – 32.5%), 10.1% (9.2 – 11%), 25.0% (23.7% – 26.4%), 44.5% (12.9 – 46.0%) and 7.6% (6.8 – 8.4%) for respondents with LBP with leg involvement respectively.

Overall effect size

The overall intervention effect size for LBP without leg involvement was small and not statistically significant (-0.17 [-0.35, 0.01]). For LBP with leg involvement, the overall intervention effect size was larger and significant -0.29 (-0.56, -0.01). For the FUOT scenario in which utilization of NSAIDs and behavioral, cognitive, and physical therapies was assumed to be 100% for individuals with LBP without leg involvement, the overall intervention effect size increased to -0.51 (-0.99, -0.04). For LBP with leg involvement utilization of NSAIDs and behavioral, cognitive, and physical therapies were also assumed to be 100% under the FUOT scenario, while utilization of surgical interventions was assumed to be 10% as surgical interventions are only recommended for a subset of individual with disc problems. Under this scenario the overall intervention effect size was -0.55 (-1.06, -0.06) for LBP with leg involvement.

LBP burden

The average DW per case for LBP in the MEPS sample was 0.093 (0.092, 0.94) for LBP without leg involvement and 0.153 (0.150, 0.155) for LBP with leg involvement, resulting in an overall weighted average DW per LBP case of 0.106 (0.105, 0.107). Figure 12 shows the distributions of LBP-specific DWs in the MEPS sample for the intervention-deleted and FUOT scenarios compared to the estimated current DW distribution. After adjusting DWs to remove the effect of healthcare interventions, the average DW per case was 0.115 (0.093, 0.134) for LBP without leg involvement and 0.220 (0.159, 0.284) for LBP

with leg involvement resulting in an overall weighted average DW of 0.138 (0.107, 0.170). Under the FUOT scenario, the average DWs for cases of LBP without and with leg involvement were 0.063 (0.028, 0.111) and 0.112 (0.041, 0.212) respectively, for a weighted average DW per LBP case of 0.073 (0.031, 0.131).

The HAQI value for the US in 2007 was 87.61. Assuming a linear relationship between HAQI and the average LBP DW per case, it was estimated that a 10-unit increase in HAQI value leads to a -0.0036-unit (-0.0072 - -0.00015) reduction in average LBP DW per case. Iceland had the highest HAQI value in 2017 at 97.17 which was used as the highest HAQI value to calculate avoidable burden. Based on this relationship, the maximum possible proportion of burden avoided was 24.4% (1.4-41.3%). For each GBD super region, Figure 13 shows the proportion of total LBP YLDs that have been averted, the proportion that could be avoided through gains in HAQI alone (avoidable) and the proportion that could be avoided through increased utilization of the optimal treatment (avoidable – FUOT) as well as the remaining burden that is not affected by increases in HAQI or intervention utilization. The high-income super region was estimated to have the highest proportion of averted burden at 22.8% (1.3-38.5%) while sub-Saharan African had the lowest averted burden at 8.9% (0.5-15.1%). Realization of the FUOT scenario was estimated to result in a 22.6% (4.1-31.2%) reduction in LBP YLDs. After accounting for the burden that could be avoided through improvements in healthcare access and quality and increases in intervention utilization, 53.0% (28.0-95.4%) of LBP burden remains.

Discussion

Health interventions have the potential to reduce LBP burden. Surgical interventions and NSAIDs elicit small but significant reductions in LBP disability while behavioral, cognitive, and physical therapies also have a small but non-significant effect. In high-income settings with a high degree of personal health system access and quality, existing interventions for LBP are estimated to have reduced burden by nearly one quarter (22.8%). The impact of healthcare interventions could be almost doubled by full utilization of the optimal set of interventions (additional 22.6% reduction in burden), though achieving 100% utilization of interventions for LBP represents an aspirational goal that is rarely achieved in practice. However, even in the unlikely scenario of full coverage, 53% of LBP burden remains. The large proportion of LBP burden that currently cannot be addressed by existing healthcare technologies highlights the role of research and development in creating novel approaches to treat low back pain.

The ability to incorporate sources of information on the effect of interventions on severity and to estimate the association between HAQI and average disability represents a novel approach to estimation of condition severity within the GBD. The large uncertainty intervals for estimates of the proportion of burden avoided through use of interventions for LBP reflect the large degree of uncertainty associated with the effect sizes for each intervention group. The method to translate intervention effects into averted and avoidable burden also accounts for uncertainty in GBD estimates of LBP prevalence, HAQI values, and the distribution of individual comorbidity-corrected LBP disability weights from the MEPs sample. In capturing each of these sources of variation, this method better characterizes estimates of average disability across geography. Importantly, uncertainty in health-state specific DWs and the translation between individual SF-12 score and individual cumulative DW for MEPS respondents was not accounted for in this analysis as the translation function developed for the GBD severity analysis does not account for uncertainty in this relationship. The low precision of avoided and avertable burden estimates reflects the limited amount of information available to characterize the relationship between intervention effect, utilization, and LBP severity and highlights the shortcomings of the current GBD assumption of constant severity across geographies.

Literature on healthcare interventions for LBP is heterogeneous and difficult to standardize. The intervention classes used in this study were defined to represent clinically meaningful groupings. Still, classes were subjectively constructed, with studies being assigned to the most appropriate class based on author's discretion. A substantial degree of heterogeneity in effect size remains among interventions

within the same class. The behavioral, cognitive, and behavioral therapy class in particular contains a number of interventions that range from involved multidisciplinary biopsychosocial rehabilitation to more straightforward mobilization exercises. This class was not further divided into smaller grouping due to the large degree of overlap between study-specific interventions. While some studies focused on a single domain, most interventions incorporated physical, cognitive, and behavioral dimensions. Given this substantial degree of crossover and the limited information provided by Cochrane reviews on intervention protocol, it was determined that it would not be feasible to further distinguish between studies in this class in any meaningful way. If more information on specific activities carried out for each component study were to be extracted from the original source, estimation of effect size and utilization for more granular intervention groupings could provide a more accurate estimation of the overall intervention effect.

Additionally, trials catalogued in Cochrane reviews varied across a number of attributes including eligibility criteria, LBP case definitions, follow-up duration, blinding protocol, and outcomes assessed. Each of these characteristics might be expected to result in differential ascertainment of intervention effect for the same intervention across different studies. At present, no bias covariates have been included in meta-regression models used to pool effect sizes within intervention classes. Including such covariates (e.g. publication year, follow-up duration) may improve estimates of between study heterogeneity and increase the precision of effect estimates for each intervention class.

Use of the standardized mean difference as a measure of effect size also has several limitations. Because this measure relies on changes in the standard deviation of an outcome measure, pooling of standardized mean differences across studies requires the assumption that changes in the distribution of outcome measures represent true differences rather than expected variation between study populations.[14] Additionally, use of the standardized mean difference to obtain a single effect size across various outcome measures requires the assumption that these measures are all linear transformations of one another. Although not empirically validated across all outcomes measures used in LBP trials, this assumption is generally accepted in meta-analysis literature.[15] The Roland Morris Disability Questionnaire and the Oswestry Low Back Disability Questionnaire were the two instruments most commonly utilized to measure LBP-associated disability. Studies comparing these instruments showed strong and significant correlation the two scales.[16,17] Additionally, it was assumed that the standardized mean difference calculated from intervention trials could be applied to SF-12 measurements. One study comparing the Roland Morris and Oswestry questionnaires found moderate correlation between these scales and the 36-item Short Form Health Survey.[17] The appropriateness of this assumption could be further tested by comparing the standardized mean differences for SF-12 and the Roland Morris or Oswestry Questionnaires from trials measuring both quality of life and LBP-specific disability.

It was also assumed that the effect of each intervention class is independent of all other classes though this may not be true for each combination of effects. Further analyses should be conducted to determine if the effect modification exists between any of the intervention classes. Evidence of effect modification would also require more detailed information on the overlap between utilization of each intervention class, which is not currently available from the MEPS dataset.

Despite these limitations the above approach can be generalized to other diseases and conditions for which severity is not well characterized in epidemiological literature in order to produce more plausible estimates of disease burden. Further applications of this work could include an analysis of the relationship between HAQI and LBP severity over time to better characterize the relationship between these two variables. Additionally, coupling this work with economic analyses of intervention packages could be used to prioritize funding and budget allocation in resource constrained settings. Through a comprehensive effort to review and synthesize all available evidence on the effect of interventions of condition severity, it is possible to understand the degree to which investment in such interventions is able to avert disease burden and improve health globally.

Implications

By leveraging information on the efficacy of healthcare interventions, the above approach represents a substantial improvement over the current GBD methodology which assumes constant severity across location. GBD currently has no avenue to incorporate the wealth of information on healthcare intervention effects and wholly ignores the effect of treatment for many of the leading causes of disability. The improved approach presented herein not only allows for the quantification of the effect of existing medical technologies on condition severity, but it also allows these estimates to be translated into measures of averted, avoidable, and remaining burden. Such quantities should be of great importance to policy makers and funders looking to expand coverage of healthcare interventions, generate new intervention research, and strengthen health system performance.

References

1. Murray, C. J. L. & Lopez, A. D. Measuring the Global Burden of Disease. *N. Engl. J. Med.* **369**, 448–457 (2013).
2. Salomon, J. A. *et al.* Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet* **380**, 2129–2143 (2012).
3. Burstein, R. *et al.* Estimating distributions of health state severity for the global burden of disease study. *Popul. Health Metr.* **13**, 31 (2015).
4. Fullman, N. *et al.* Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: a systematic analysis from the Global Burden of Disease Study 2016. *The Lancet* **391**, 2236–2271 (2018).
5. James, S. L. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1789–1858 (2018).
6. Medical Expenditure Panel Survey. Available at: https://meps.ahrq.gov/mepsweb/. (Accessed: 1st May 2019)
7. Cochrane Reviews | Cochrane Library. Available at: https://www.cochranelibrary.com/. (Accessed: 30th April 2019)
8. Chou, R. *et al.* Diagnosis and Treatment of Low Back Pain: A Joint Clinical Practice Guideline from the American College of Physicians and the American Pain Society. *Ann. Intern. Med.* **147**, 478 (2007).
9. Chou, R. *et al.* Surgery for low back pain: a review of the evidence for an American Pain Society Clinical Practice Guideline. *Spine* **34**, 1094–1109 (2009).

10. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2019).

11. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).

12. Lumley, T. Network meta-analysis for indirect treatment comparisons. *Stat. Med.* **21**, 2313–2324 (2002).

13. Faraone, S. V. Interpreting Estimates of Treatment Effects. *Pharm. Ther.* **33**, 700–711 (2008).

14. Cochrane Handbook for Systematic Reviews of Interventions. Available at: https://training.cochrane.org/handbook. (Accessed: 6th May 2019)

15. Effect Sizes Based on Means. in *Introduction to Meta-Analysis* 21–32 (John Wiley & Sons, Ltd, 2009). doi:10.1002/9780470743386.ch4

16. van Hooff, M. L., Spruit, M., Fairbank, J. C. T., van Limbeek, J. & Jacobs, W. C. H. The Oswestry Disability Index (version 2.1a): validation of a Dutch language version. *Spine* **40**, E83-90 (2015).

17. Monticone, M. *et al.* Responsiveness of the Oswestry Disability Index and the Roland Morris Disability Questionnaire in Italian subjects with sub-acute and chronic low back pain. *Eur. Spine J.* **21**, 122–129 (2012).

Included Cochrane Reviews

Chaparro L, Furlan A, Deshpande A, Mailis-Gagnon A, Atlas S, Turk D. Opioids compared to placebo or other treatments for chronic low-back pain. *Cochrane Database of Systematic Reviews* 2013. DOI:10.1002/14651858.CD004959.pub4.

Dagenais S, Yelland M, Del Mar C, Schoene M. Prolotherapy injections for chronic low-back pain. *Cochrane Database of Systematic Reviews* 2007. DOI:10.1002/14651858.CD004059.pub3.

Dahm K, Brurberg K, Jamtvedt G, Hagen K. Advice to rest in bed versus advice to stay active for acute low-back pain and sciatica. *Cochrane Database of Systematic Reviews* 2010. DOI:10.1002/14651858.CD007612.pub2.

Ebadi S, Henschke N, Nakhostin Ansari N, Fallah E, van Tulder M. Therapeutic ultrasound for chronic low-back pain. *Cochrane Database of Systematic Reviews* 2014. DOI:10.1002/14651858.CD009169.pub2.

Enthoven W, Roelofs P, Deyo R, van Tulder M, Koes B. Non-steroidal anti-inflammatory drugs for chronic low back pain. *Cochrane Database of Systematic Reviews* 2016. DOI:10.1002/14651858.CD012087.

Franke H, Fryer G, Ostelo R, Kamper S. Muscle energy technique for non-specific low-back pain. *Cochrane Database of Systematic Reviews* 2015. DOI:10.1002/14651858.CD009852.pub2.

French S, Cameron M, Walker B, Reggars J, Esterman A. Superficial heat or cold for low back pain. *Cochrane Database of Systematic Reviews* 2006. DOI:10.1002/14651858.CD004750.pub2.

Furlan A, Giraldo M, Baskwill A, Irvin E, Imamura M. Massage for low-back pain. *Cochrane Database of Systematic Reviews* 2015. DOI:10.1002/14651858.CD001929.pub3.

Furlan A, van Tulder M, Cherkin D, *et al.* Acupuncture and dry-needling for low back pain. *Cochrane Database of Systematic Reviews* 2005. DOI:10.1002/14651858.CD001351.pub2.

Gibson J, Waddell G. Surgery for degenerative lumbar spondylosis. *Cochrane Database of Systematic Reviews* 2005. DOI:10.1002/14651858.CD001352.pub3.

Hayden J, van Tulder M, Malmivaara A, Koes B. Exercise therapy for treatment of non-specific low back pain. *Cochrane Database of Systematic Reviews* 2005. DOI:10.1002/14651858.CD000335.pub2.

Henschke N, Ostelo R, van Tulder M, *et al.* Behavioural treatment for chronic low-back pain. *Cochrane Database of Systematic Reviews* 2010. DOI:10.1002/14651858.CD002014.pub3.

Kamper S, Apeldoorn A, Chiarotto A, *et al.* Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. *Cochrane Database of Systematic Reviews* 2014. DOI:10.1002/14651858.CD000963.pub3.

Khadilkar A, Odebiyi D, Brosseau L, Wells G. Transcutaneous electrical nerve stimulation (TENS) versus placebo for chronic low-back pain. *Cochrane Database of Systematic Reviews* 2008. DOI:10.1002/14651858.CD003008.pub3.

Maas E, Ostelo R, Niemisto L, *et al.* Radiofrequency denervation for chronic low back pain. *Cochrane Database of Systematic Reviews* 2015. DOI:10.1002/14651858.CD008572.pub2.

Macedo L, Saragiotto B, Yamato T, *et al.* Motor control exercise for acute non-specific low back pain. *Cochrane Database of Systematic Reviews* 2016. DOI:10.1002/14651858.CD012085.

Machado G, Ferreira P, Yoo R, *et al.* Surgical options for lumbar spinal stenosis. *Cochrane Database of Systematic Reviews* 2016. DOI:10.1002/14651858.CD012421.

Marin T, Van Eerd D, Irvin E, *et al.* Multidisciplinary biopsychosocial rehabilitation for subacute low back pain. *Cochrane Database of Systematic Reviews* 2017. DOI:10.1002/14651858.CD002193.pub2.

Overdevest G, Jacobs W, Vleggeert-Lankamp C, Thomé C, Gunzburg R, Peul W. Effectiveness of posterior decompression techniques compared with conventional laminectomy for lumbar stenosis. *Cochrane Database of Systematic Reviews* 2015. DOI:10.1002/14651858.CD010036.pub2.

Parreira P, Heymans M, van Tulder M, *et al.* Back Schools for chronic non-specific low back pain. *Cochrane Database of Systematic Reviews* 2017. DOI:10.1002/14651858.CD011674.pub2.

Rubinstein S, Terwee C, Assendelft W, de Boer M, van Tulder M. Spinal manipulative therapy for acute low-back pain. *Cochrane Database of Systematic Reviews* 2012. DOI:10.1002/14651858.CD008880.pub2.

Rubinstein S, van Middelkoop M, Assendelft W, de Boer M, van Tulder M. Spinal manipulative therapy for chronic low-back pain. *Cochrane Database of Systematic Reviews* 2011. DOI:10.1002/14651858.CD008112.pub2.

Saragiotto B, Machado G, Ferreira M, Pinheiro M, Abdel Shaheed C, Maher C. Paracetamol for low back pain. *Cochrane Database of Systematic Reviews* 2016. DOI:10.1002/14651858.CD012230.

Saragiotto B, Maher C, Yamato T, *et al.* Motor control exercise for chronic non-specific low-back pain. *Cochrane Database of Systematic Reviews* 2016. DOI:10.1002/14651858.CD012004.

Urquhart D, Hoving J, Assendelft W, Roland M, van Tulder M. Antidepressants for non-specific low back pain. *Cochrane Database of Systematic Reviews* 2008. DOI:10.1002/14651858.CD001703.pub3.

Walker B, French S, Grant W, Green S. Combined chiropractic interventions for low-back pain. *Cochrane Database of Systematic Reviews* 2010. DOI:10.1002/14651858.CD005427.pub2.

Wegner I, Widyahening I, van Tulder M, *et al.* Traction for low-back pain with or without sciatica. *Cochrane Database of Systematic Reviews* 2013. DOI:10.1002/14651858.CD003010.pub5.

Wieland L, Skoetz N, Pilkington K, Vempati R, D'Adamo C, Berman B. Yoga treatment for chronic non-specific low back pain. *Cochrane Database of Systematic Reviews* 2017. DOI:10.1002/14651858.CD010671.pub2.

Yamato T, Maher C, Saragiotto B, *et al.* Pilates for low back pain. *Cochrane Database of Systematic Reviews* 2015. DOI:10.1002/14651858.CD010265.pub2.

Yousefi-Nooraie R, Schonstein E, Heidari K, *et al.* Low level laser therapy for nonspecific low-back pain. *Cochrane Database of Systematic Reviews* 2008. DOI:10.1002/14651858.CD005107.pub4.

Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. *Cochrane Database of Systematic Reviews* 2016. DOI:10.1002/14651858.CD010264.pub2.

Tables and Figures

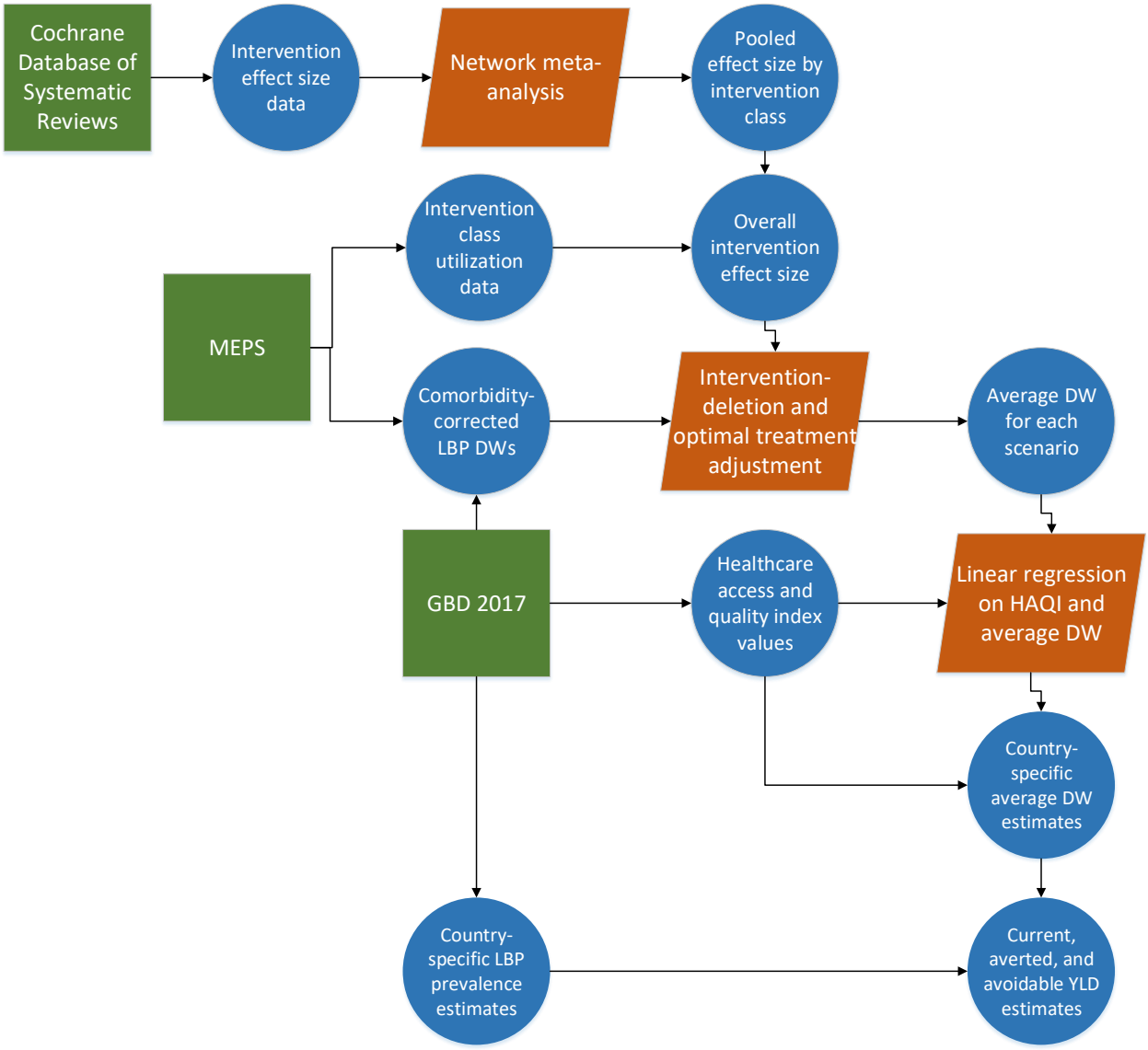

Figure 1 - ***Methodological flowchart*** *High-level overview of study methodology*

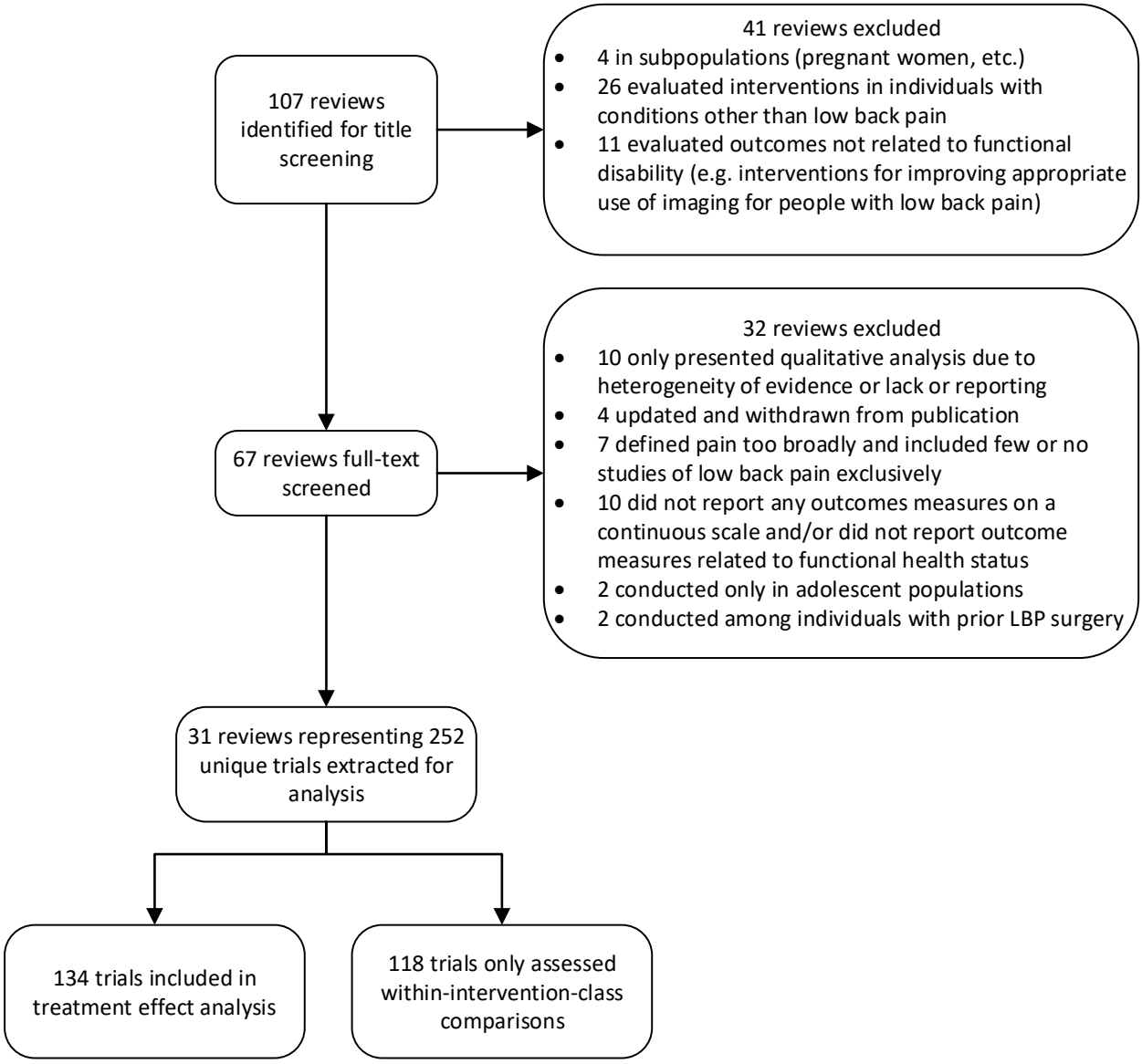

Figure 2- ***Data extraction flow diagram*** *Source counts and exclusion criteria for studies returned from the Cochrane Database of Systematic Reviews.*

Table 1 – ***Network meta-analysis comparison counts*** *Number of trials for each intervention-comparison pair in treatment effect network meta-analysis*

Intervention class	Comparison class	Number of trials
Behavioral, cognitive, and physical therapies	Placebo or usual care (reference)	126
Behavioral, cognitive, and physical therapies	NSAIDs	4
Behavioral, cognitive, and physical therapies	Surgical interventions	3
Behavioral, cognitive, and physical therapies	Non-opioid, non-NSAID analgesics	1
NSAIDs	Placebo or usual care (reference)	4
Non-opioid, non-NSAID analgesics	Placebo or usual care (reference)	2
Opioid analgesics	Placebo or usual care (reference)	11
Surgical interventions	Placebo or usual care (reference)	9

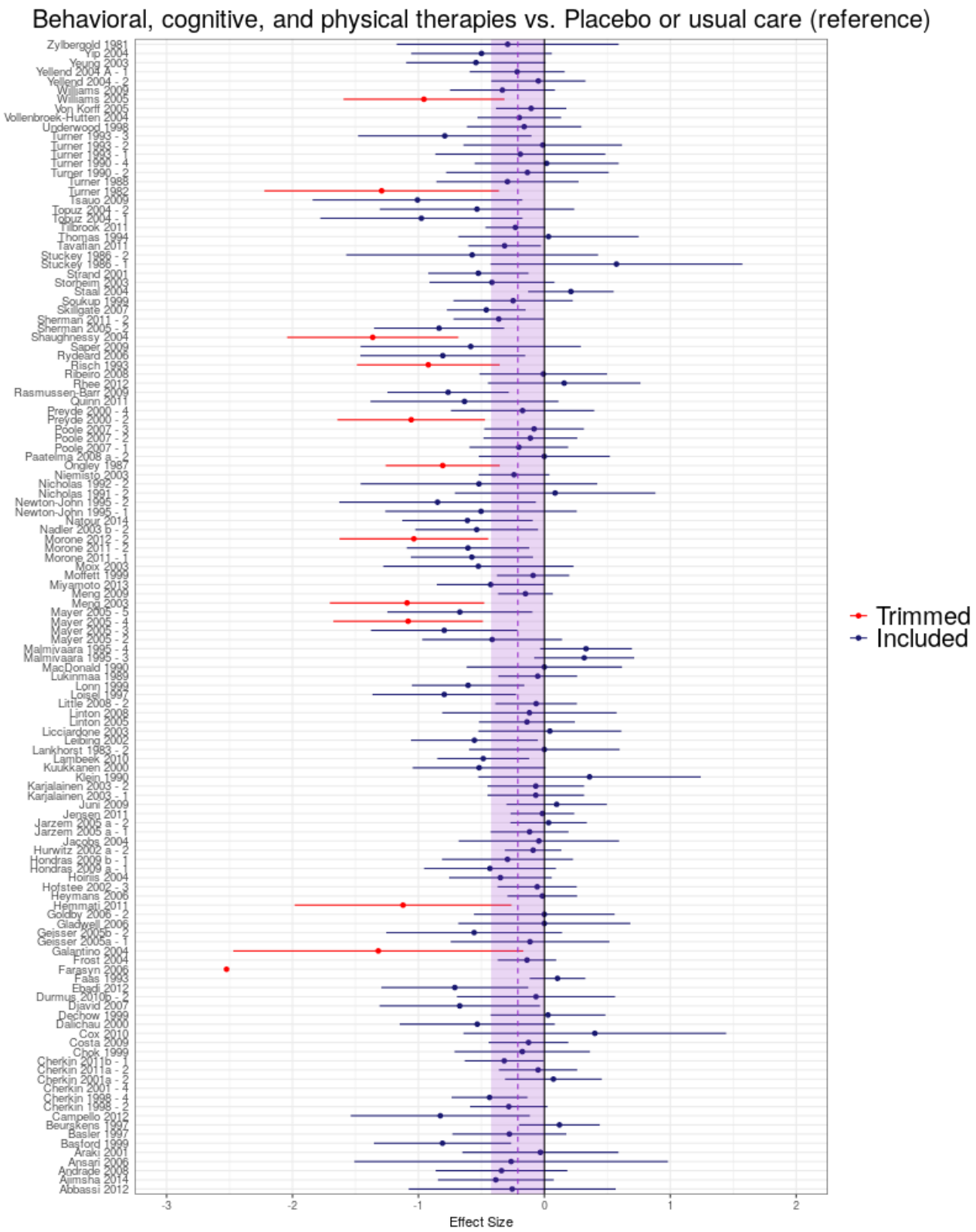

Figure 3 – ***Forest plot of behavioral cognitive and physical therapies vs reference*** *Standardized mean differences and SD for studies comparing behavioral, cognitive, and physical therapies to placebo or usual care. Estimated pooled standardized mean difference is shown as the purple dashed line with SD shaded in purple*

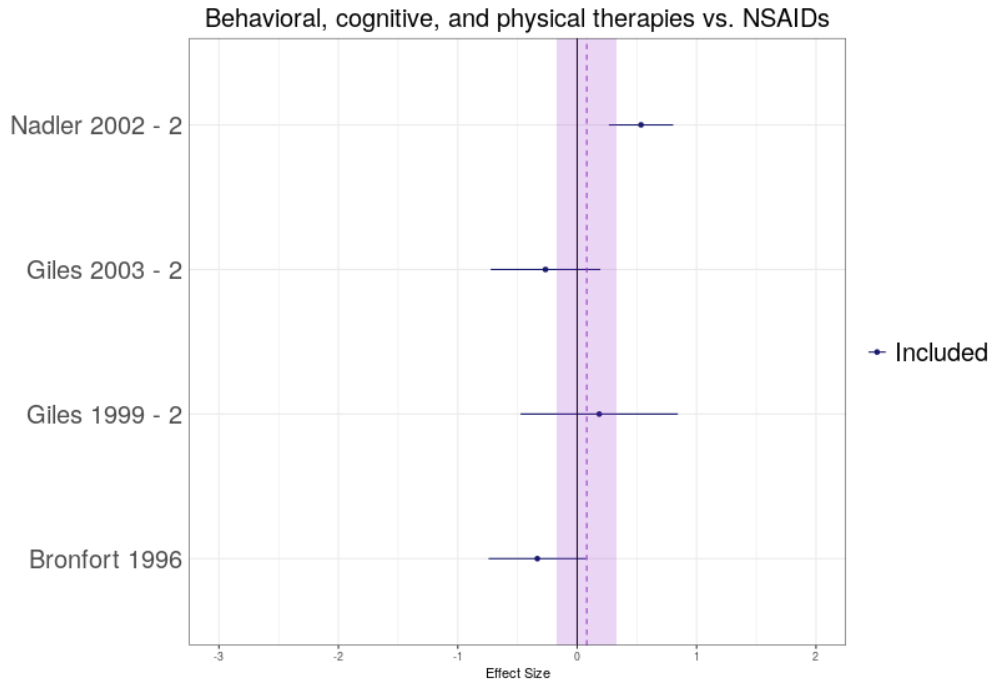

Figure 4 – ***Forest plot of behavioral cognitive and physical therapies vs NSAIDs*** *Standardized mean differences and SDs for studies comparing behavioral, cognitive, and physical therapies to NSAIDs. Estimated pooled standardized mean difference is shown as the purple dashed line with SD shaded in purple*

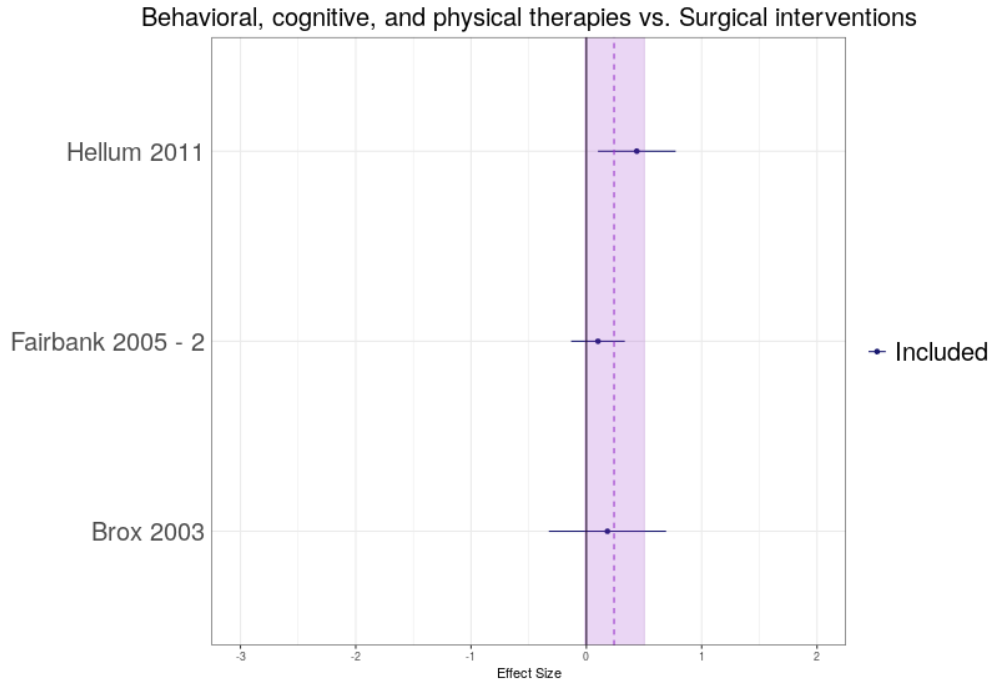

Figure 5 – ***Forest plot of behavioral cognitive and physical therapies vs surgical interventions*** *Standardized mean differences and SDs for studies comparing behavioral, cognitive, and physical therapies to surgical interventions. Estimated pooled standardized mean difference is shown as the purple dashed line with SD shaded in purple*

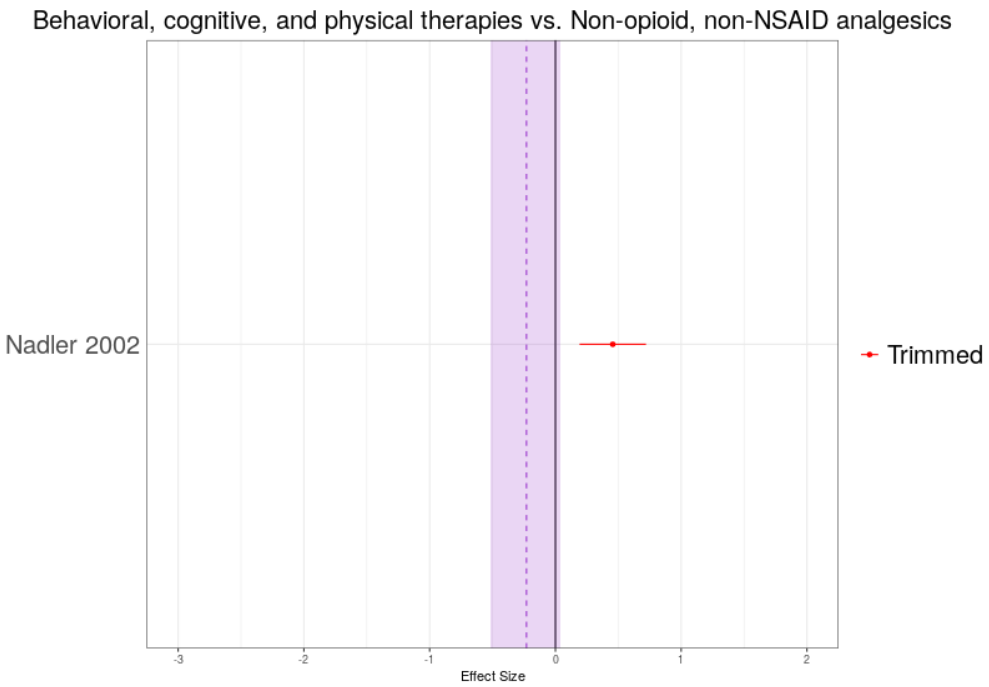

Figure 6 – ***Forest plot of behavioral cognitive and physical therapies vs non-opioid non-NSAID analgesics*** *Standardized mean differences and SDs for studies comparing behavioral, cognitive, and physical therapies to non-opioid, non-NSAID analgesics. Estimated pooled standardized mean difference is shown as the purple dashed line with SD shaded in purple*

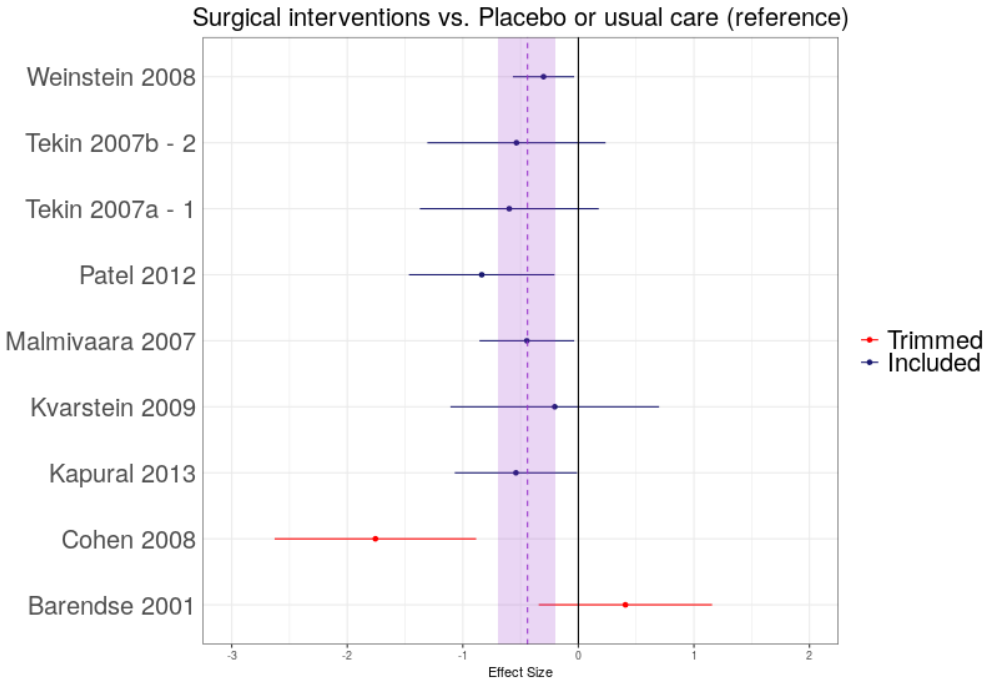

Figure 7 – ***Forest plot of surgical interventions vs reference*** *Standardized mean differences and SDs for studies comparing surgical interventions to placebo or usual care. Estimated pooled standardized mean difference is shown as the purple dashed line with SD shaded in purple*

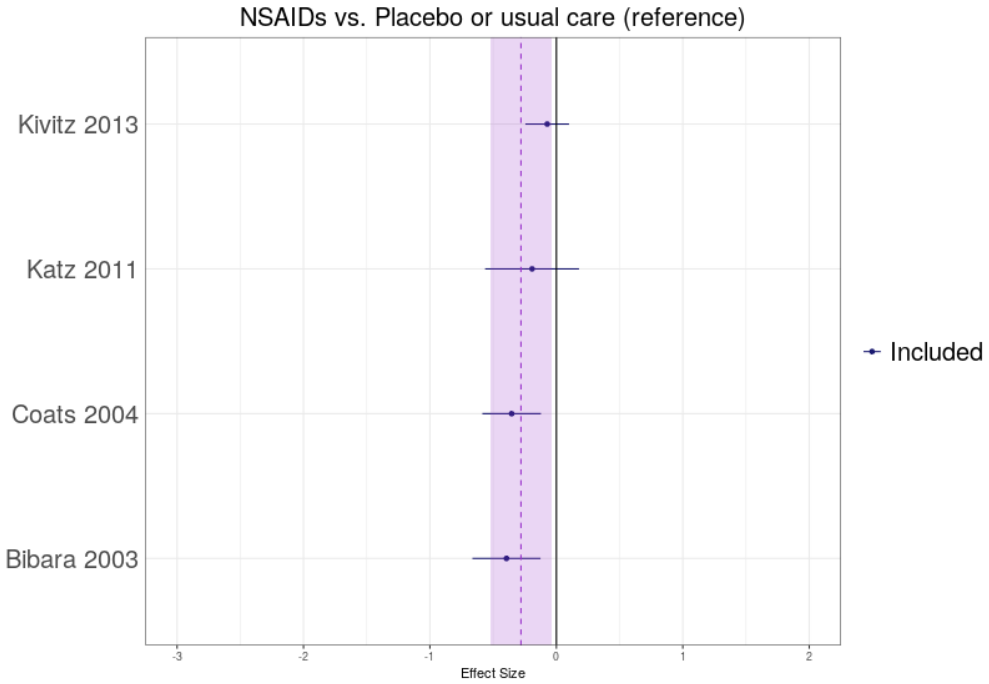

*Figure 8 – **Forest plot of NSAIDs vs reference** Standardized mean differences and SDs for studies comparing NSAIDs to placebo or usual care. Estimated pooled standardized mean difference is shown as the purple dashed line with SD shaded in purple*

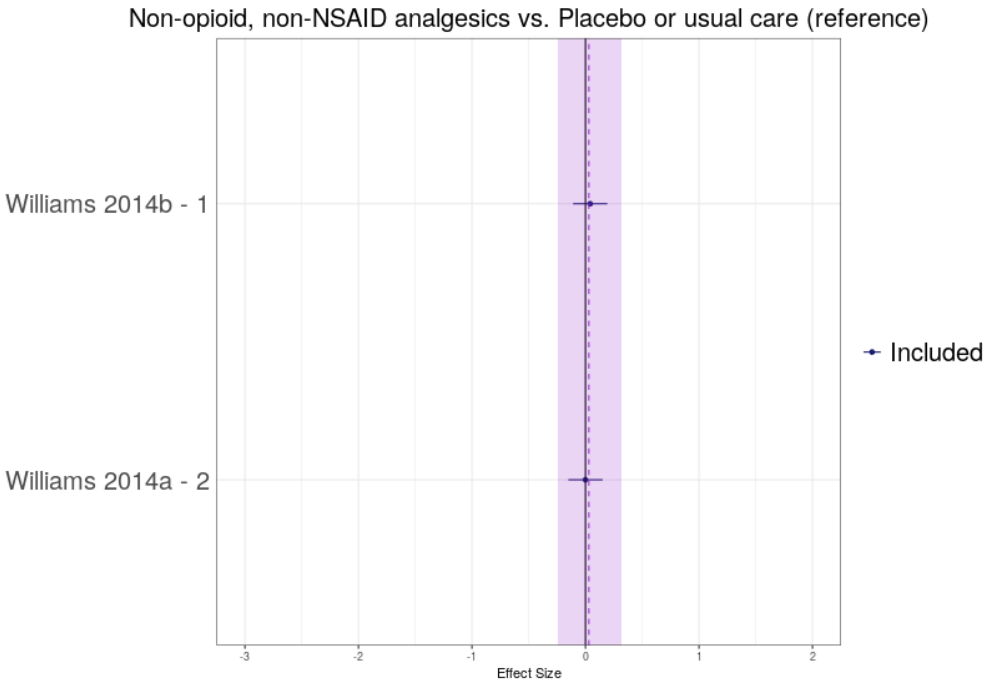

*Figure 9 – **Forest plot of non-opioid non-NSAID analgesics vs reference** Standardized mean differences and SDs for studies comparing non-opioid, non-NSAID analgesics to placebo or usual care. Estimated pooled standardized mean difference is shown as the purple dashed line with SD shaded in purple*

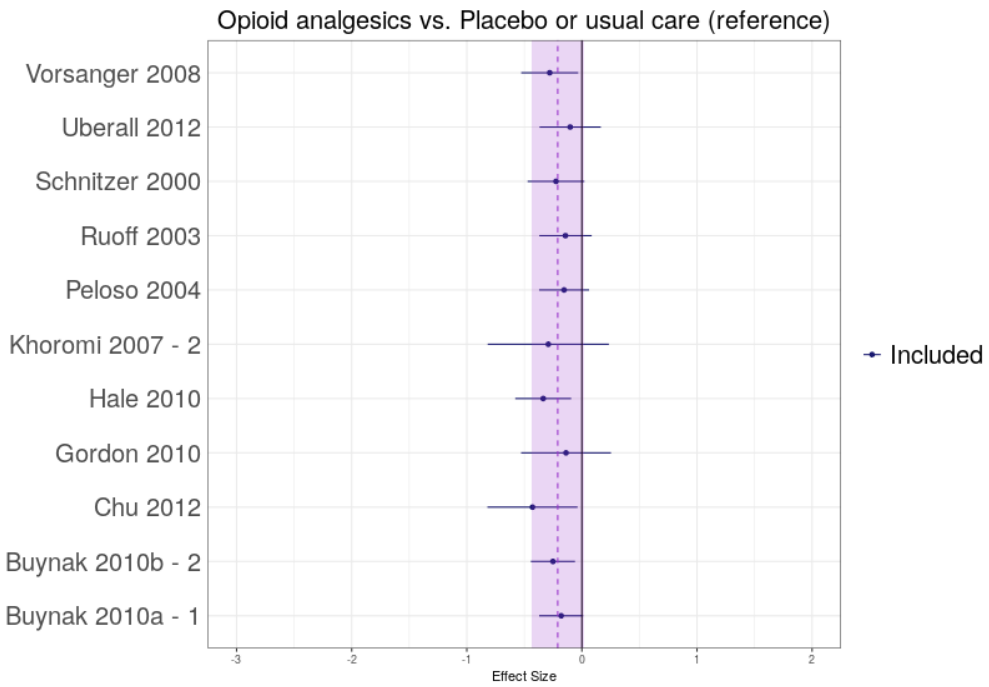

*Figure 10 – **Forest plot of opioid analgesics vs reference** Standardized mean differences and SDs for studies comparing opioid analgesics to placebo or usual care. Estimated pooled standardized mean difference is shown as the purple dashed line with SD shaded in purple*

*Table 2 – **Intervention class effect size estimates** Estimated standardized mean differences compared to placebo and corresponding confidence intervals with and without adjustment for estimated between study heterogeneity for each intervention class*

		Unadjusted confidence interval		Adjusted confidence interval	
Intervention class	**SMD**	**Lower**	**Upper**	**Lower**	**Upper**
Surgical interventions	-0.44	-0.60	-0.30	-0.72	-0.18
NSAIDs	-0.28	-0.42	-0.15	-0.52	-0.04
Opioid analgesics	-0.22	-0.32	-0.12	-0.44	0.01
Behavioral, cognitive, and physical therapies	-0.21	-0.26	-0.17	-0.43	0.01
Non-opioid, non-NSAID analgesics	0.02	-0.16	0.20	-0.25	0.29

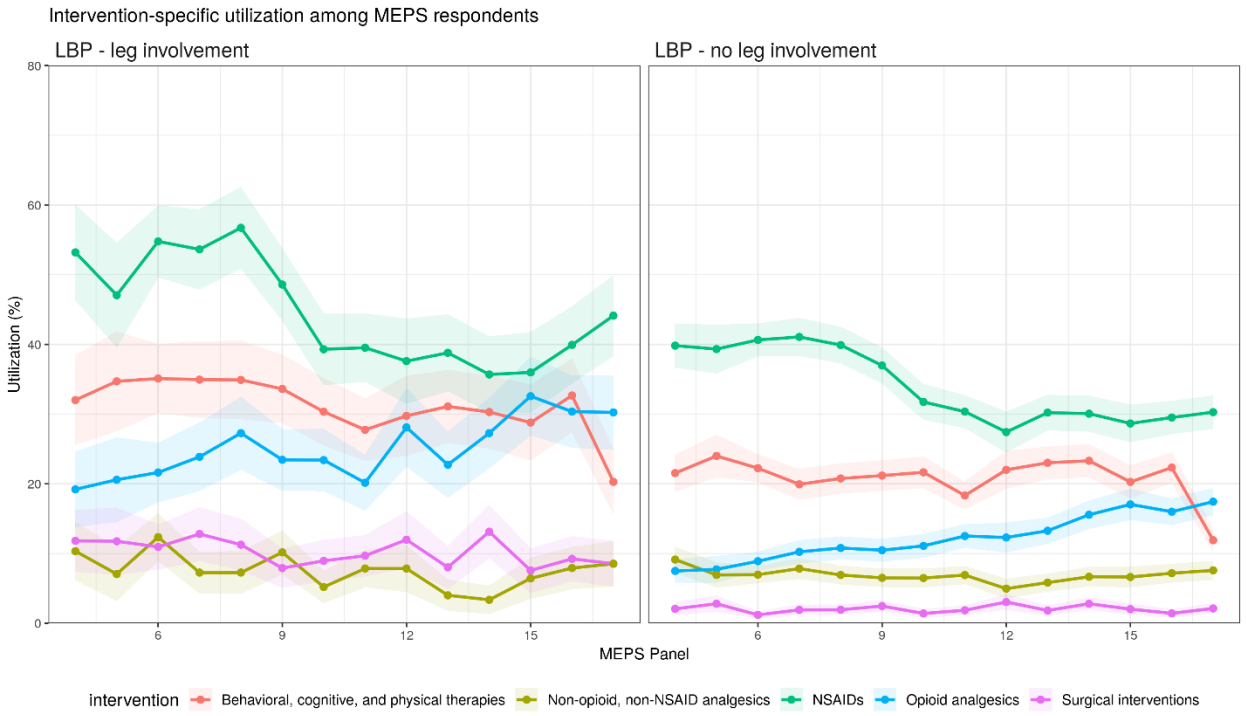

*Figure 11 – **Intervention class utilization estimates** Intervention utilization estimates among MEPS respondents by MEPS panel. Shaded areas represent 95% confidence interval*

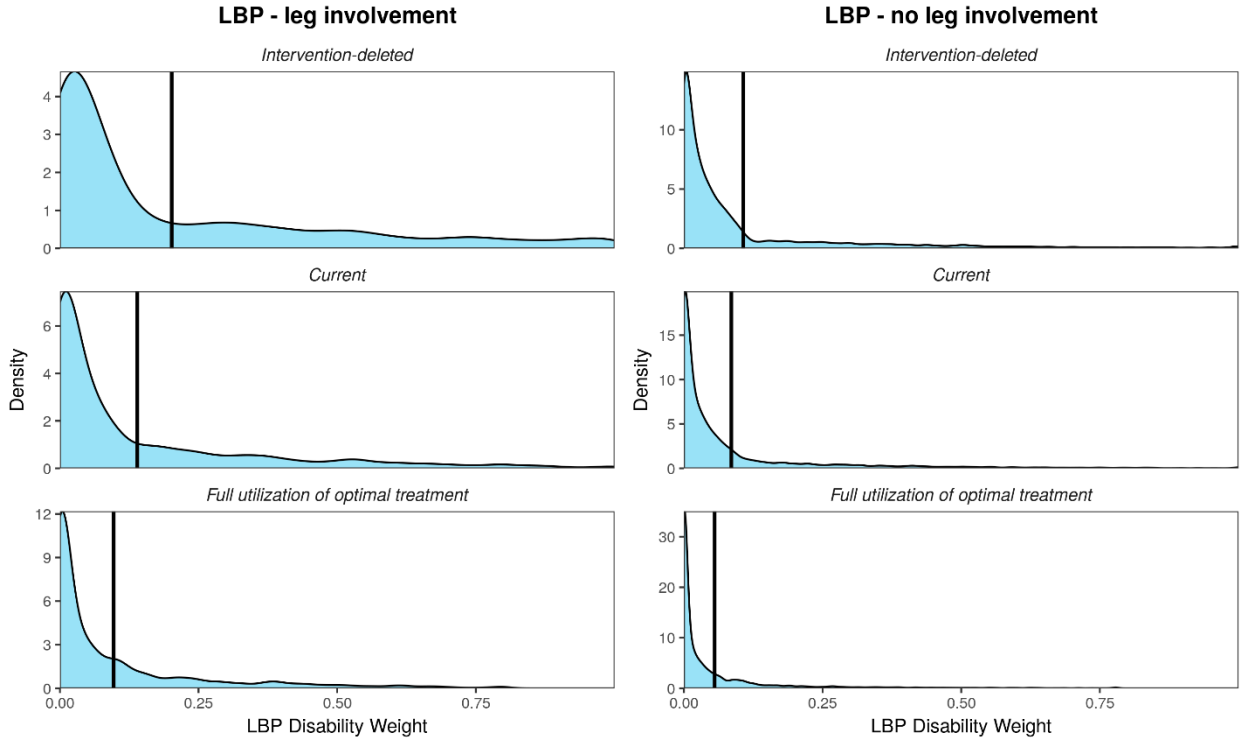

*Figure 12 – **MEPS disability weigh distributions by scenario** Density plots of DW distributions for the intervention-deleted and FUOT scenarios as compared to the current DW distribution in the MEPS sample for LBP with and without leg involvement. Black lines indicate distribution mean.*

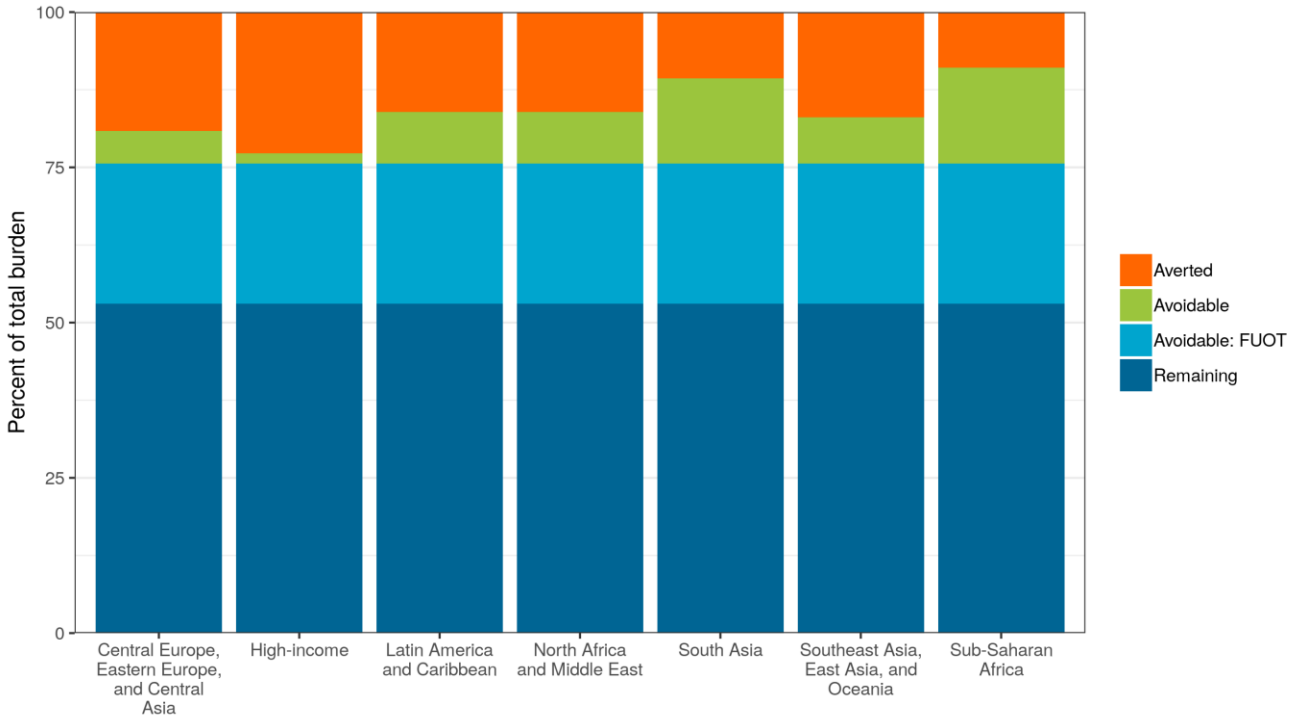

Figure 13 – ***Proportion of burden addressed by healthcare interventions*** *Averted, avoidable, and remaining LBP burden as a proportion of total LBP YLDs by GBD super region*