

APPENDIX A. SEARCH STRATEGY



HIP & KNEE REPLACEMENT – DISPARITIES

DATABASE SEARCHED:

PubMed

TIME PERIOD COVERED: 1966-2010

LANGUAGE:

ENGLISH

SEARCH STRATEGY:

arthroplasty, replacement, hip OR arthroplasty, replacement, knee OR arthritis/surgery OR osteoarthritis, knee/surgery OR osteoarthritis, hip/surgery OR (hip AND surger*)(OR (knee AND surger*) OR (joint AND replac*) OR (knee AND replac*) OR (hip AND replac*) OR (joint AND arthroplast*) OR (hip AND arthroplast*) OR (knee AND arthroplast*)

AND

minority groups OR african americans OR hispanic americans OR african continental ancestry group OR racial OR ethnic OR minorities OR gender OR sex OR age[ti] OR age distribution

AND

disparity OR disparities OR difference* OR variation*

APPENDIX B. STUDY SELECTION FORM

Article ID:	Reviewer:
Authors:	
Title:	

1. What type of joint replacement is discussed?
(Check all that apply)

Total Knee Replacement (TKR)

Total Hip Replacement (THR)

Other: (specify _____) ...

None

If None → Stop

2. What type of disparity is discussed?
(Check all that apply)

Racial/ethnic.....

Gender.....

Urban/rural.....

Regional.....

Socioeconomic status.....

Other: (specify _____) ...

Not clear, needs further review

None

If None → Stop

3. Which generation of study does this best fit?
(See explanation below for reference)

First

Second

Third

5. Are veterans discussed, either as the main focus or as a sub-category?

Main focus.....

Sub-category

No mention.....

6. What is the study design?

Descriptive/observational.....

Experimental

Qualitative.....

Systematic review or Meta-Analysis

Other.....

7. What is the approximate sample size?

<100

100-500

>500

8. Study origin?

Unclear.....

US.....

Non-US

Specify: _____

9. Mark as Background Article

- **First-Generation** studies document the *existence and the magnitude* of the disparities.
- **Second-generation** studies examine the *reasons for observed disparities* and could be classified as: a) patient-level factors (treatment preferences, patterns of self-care, etc.); b) provider-level factors (physician-patient communication, etc.); and c) system level factors (access to specialist care, etc.).
- **Third-generation** studies examine *interventions* to address the observed disparities.

APPENDIX C. DATA EXTRACTION FORM

ID:	Reviewer:
Author:	
Title:	

1. Is the study based solely in the US?
 Yes No STOP
2. Does the article examine an intervention to address the
 Yes
 If Yes, specify: _____
 No
3. What is the data source for the study? (check all that apply)
 a. Single institution
 b. Multi-institutional (Regional)
 c. Multi-institutional (National)
 d. VA (single or multi-VA)
 e. Medicare
 f. Medicaid
 g. NIS
 h. Other (Specify: _____)
4. Who are the primary study subjects? (check all)
 a. Patients
 b. Providers
 c. Other
5. How were patients selected?
 a. Population-based/systematic/representative sample
 b. Consecutive patients
 c. Convenience/non-representative sample
 d. Combination of above
 e. Unclear/unknown
6. Years of data collection covered*: _____
7. What is the study design?
 a. Cross-sectional
 b. Cohort/Case-control
 c. Experimental
 d. Systematic Review
 e. Unclear/unknown
 f. Background STOP
8. What is/are the Primary Outcome Measure(s)?
 a. Receipt of procedure(s)
 b. Recommendation for procedure(s)
 c. Outcome of Procedure
 d. Appropriateness for Procedure
 e. Perception of Need for Surgery
 f. Willingness to Consider Surgery
 g. Outcome Expectations of Surgery
 h. Other
 Specify: _____
 Specify: _____
9. What is/are the Secondary Outcome Measure(s)?
 a. Receipt of procedure(s)
 b. Recommendation For procedure(s)
 c. Outcome of Procedure
 d. Appropriateness for Procedure
 e. Perception of Need for Surgery
 f. Willingness to Consider Surgery
 g. Outcome Expectations of Surgery
 h. Not Applicable
 i. Other
 Specify: _____
 Specify: _____
 Specify: _____

10. **For Race:** How was race categorized?
- a. White b. Black
c. Hispanic d. Asian
e. Non-black f. Non-white
g. None
11. **For Race:** How was race determined?
- a. Administrative data b. Medical record review
c. Self-reported d. Two or more sources
e. Unknown or not reported
12. Assessment of receipt of procedure
- a. Medical record b. Administrative data
c. Self report d. Not reported/unknown
e. Not applicable
13. Assessment of primary disparity outcome (function, quality of life, receipt of procedure, etc)
- a. Medical record b. Administrative data
c. Self report d. Not reported/unknown
14. Population sample size:
- a. Total _____
b. Veteran sample size _____
15. What was the mean/median age of the patients*?
- _____
16. Response Rate
- a. Number Eligible for Study

- b. Number Declining Participation

- c. Response Rate if Reported _____
- d. Unclear/unspecified
e. Not applicable
17. Adequacy of follow-up for subjects?
- a. Complete follow-up of all subjects
b. Subjects lost to follow-up
If so, # _____ or % _____ f/u
- c. Description of those lost to f/u
d. Unclear/unspecified
e. Not applicable
18. If results were adjusted, were the following covariates included?
- a. Age
b. Gender
c. Income
d. Education
e. Insurance
f. Not Applicable
19. **For Race:** How many “unknowns” were reported?
- a. Total or percent “unknown” _____
b. Not reported

*No data denoted as 999

APPENDIX D. Newcastle Ottawa Scale Criteria Used in Quality Assessment

Selection

- 1) Representativeness of the exposed cohort
- 2) Ascertainment of exposure
- 3) Demonstration that outcome of interest was not present at start of study

Comparability

- 1) Comparability of cohorts on the basis of the design or analysis

Outcome

- 1) Assessment of outcome
- 2) Was follow-up long enough for outcomes to occur
- 3) Adequacy of follow up of cohorts

APPENDIX E. PEER REVIEW COMMENTS/AUTHOR RESPONSES

Comment	Response
General Comments	
Issues such as waiting time, access to orthopedics consultations, etc. have not been clearly documented, and it is unknown if they may relate to some of the differences observed. If these barriers indeed exist, they should be addressed through quality improvement measures.	There is a new section in the limitations that addresses this.
Abbreviations not used consistently	After double-checking for inconsistencies, some remain due to the literature itself.
Missing: Hausmann, et al, Arthritis Care & Rheum 2011, p635-642	Now included, found in update search
In general the ‘summary of findings’ sections read better than the ‘data’ sections. The data sections need revision. Revisions to the data sections should be directed at reading fluency to better convey the intended message. Substantial editorial attention to writing, paragraph structure, style, grammar, and typographical errors is suggested.	Edits have been made to address this issue.
Authors may want to do a brief re-search, using the same database and keywords, for the period from January through June 2011 as a final update prior to publication.	Update is now included
I have some reservations about the level of evidence available to make any decisions of consequence. Although there is no documentation of the levels, it appears that most cited studies are at minimum a level III or IV. It would be helpful to document by the standard definitions and the numerical system to be sure everyone is on the same page. In addition, grades of recommendation for the consensus should use a standard nomenclature such as A,B,C,I, again for clarity. When levels of evidence are so low or the mass of supporting evidence is so low and recommendation can only be I, the only conclusion would be we need focused research of the highest quality and nothing more. In this current context, the supposed disparity may have no other basis than personal preference, cultural beliefs and population bias which may not be alterable.	The ESP program uses the GRADE system, which does not use the level I, II, III approach. The limitations of the evidence are reflected in the overall “quality of evidence,” and most of these are rated as low or very low.
Executive Summary	
The summary was a little vague with respect to results. Since many individuals may only access the summary, a more precise summary of results would be appropriate, such as including how many studies were available for each key question, and whether they included or not VA populations. In general, the findings are reported as ‘few studies....’ or ‘most....’. Including number of studies and participants would be useful. This is all included in the main report, but would be useful in the summary.	We have revised the executive summary to include more detail
2 nd paragraphs notes there are “disparities” in TJR use in non-VA settings. It would be useful here to mention types of disparities that are being alluded to (e.g., race, gender, ethnicity...)	This paragraph has been updated to be more specific.
page 2, Key Question #1, 2 nd paragraph notes that future research is unlikely to change confidence on the estimate of the effect. It should be qualified here (and elsewhere in the synthesis where this is also mentioned) that future research is still important for evaluating whether there are any temporal trends in disparities (e.g., do these change over time in response to any policies, interventions, etc).	Very good point, we have updated the relevant sections accordingly.

Comment	Response
On page 2 in the executive summary----it would be nice in the summary to add the number of studies contributing to the literature for each KQ	These numbers have been added at the beginning of the results section.
Background	
In the Background on page 1, page 5, and elsewhere the report refers to “disparities.” This description is too general because the report only addresses gender and racial disparities. Clarification of this usage to use language such as “gender and racial disparities” in place of “disparities” is suggested.	We revised the draft to clarify.
The authors are limited by research papers which primarily address only two racial / ethnic groups (White, and African-American) and don't clearly address educational, socioeconomic and regional effects.	This is a limitation of the primary literature. Almost all the disparity literature deals with gender and race. Even within gender disparities, VA data are very scarce. These additional potential sources of disparities are now noted in the limitations.
Methods	
Search strategy includes ‘peer-reviewed’ articles. How is this assessed? Do the authors mean original publications? If reviews were included, how was it determined if they were peer-reviewed	We revised this to indicate that anything indexed on PubMed was potentially eligible
In the Methods (page 6) under the heading ‘search strategy’ more detailed description of the search terms should be provided. At a minimum indicate the surgical procedures THR and TKR.	We have added some of our specific terms, and the entire search strategy is in Appendix A.
Flow	
Figure 3 Literature Flow seems to have a discrepancy in the number of articles categorized in the bottom row of boxes. There are $22+35+1 = 58$ studies categorized in the bottom row. The row above indicates there were 69 articles assessed. So it seems there are 11 articles (69 minus 58 = 11) that are not categorized.	Additional explanations have been added to clarify the overlapping nature of the categories, which accounts for the numerical discrepancy.
Figure 3 - It may also be useful to provide a breakdown of which or how many articles addressed racial disparities and how many addressed gender disparities. The current breakdown seems to indicate only 1 article addressed gender disparities	This is correct, there was only one gender article.
I can't follow Figure 3 and the numbers. It says 69 articles were assessed but the numbers below don't add up. Please clarify/fix. Also, I think it would help the reader to explain the literature groups below the figure.	Additional explanations have been added to clarify the overlapping nature of the categories, and the groups are now referred to by key question, rather than generation, for clarity.
Study Design	
Authors state that study design was not used as inclusion/exclusion criterion. However, Figure 3 includes inappropriate study design as a rejection criterion.	We have reworded the figure to be more specific.
Results	
The first set of results for Key Question #2 is related to a comparison of VA and non-VA county hospitals. It would be helpful to the reader if there is a clear statement about how these data relate to the key questions (e.g. differences according to a system-level factor?).	Updates have been made to address this.

Comment	Response
Page 12 – 2 nd paragraph under VA Data, 3 rd sentence – it is not completely clear what the comparison is for the OR, and there is a grammar issue in the sentence.	This sentence has been updated.
Page 12, three studies are alluded to – is the 2 nd paragraph in this section about the third study?	Yes, this has been noted in the text.
Page 12, 2 nd paragraph – which ICD-9 codes were used?	ICD-9 codes have been added where necessary
Page 13, 3 rd paragraph under Non-VA data: some discussion of the magnitude of differences in TKR rates would be helpful.	The Kane review (published in arthritis care and research), does not list actual rates of use, just that rates in one group are larger or smaller than others. We think that the rates presented in the following studies below can present a sense of magnitude of differences in rates.
Page 17, Summary of Findings – doesn't fully summarize the results (or lack thereof) regarding patient, provider, and system level factors.	Changes have been made.
In some places it is clear which cohort is being referred to, in other places it is less clear (e.g., “another VA cohort” on p18, 2 nd paragraph). Throughout, it would be helpful to have a consistent way of referring to each study in the table.	Updates have been made to better identify the cohort (e.g. the Cleveland cohort)
It is not clear that the general information under Non-VA Data that starts on p21 is needed. It seems a bit out of place here.	This section has been updated for better flow.
P24, Summary of findings – it would be helpful to compare / contrast this with VA data, mentioning any different findings or just areas in which there are may be more data for non-VA vs. VA.	Updates have been made to address this.
For KQ1 there is one ‘summary of findings’ section at the end that includes both VA and non-VA data. For KQ2, the structure differs and was confusing at first – that is, within the KQ2 sections there are three ‘summary’ sections for each of VA + non-VA, VA, and non-VA.	This structure was used due to the volume of literature in the sections.
There is only one study examining gender disparities in the VA. This finding of limited research related to gender is not highlighted in the summary of findings.	The summary has been updated to reflect this.
Note: on page 12, first sentence under VA data---I think you want a “the” before VA. Also, in the 2 nd paragraph, note there is a comma rather than a period in the pt estimate of 0.3 %	Changes have been made.
page 16---para 30--- fix tense of first sentence. Note also that the last sentence of this paragraph does not explain what the 2 fold higher odds are of????	This sentence has been removed.
Page 17. Please clarify last sentence of 2 nd paragraph	Edits have been made to address this issue.
Page 18. 2 nd paragraph, 2 nd to last sentence----I think you mean TJR rather than OA	OA has been verified.
Page 19-last sentence----take out “thus”	Fixed
Page 23----2 nd paragraph----review the middle sentence that states “social support between various racial groups after undergoing a hip fracture.....	This sentence has been reworded for grammar and clarity.
It is hard to get too excited about KQ 3 since there seemed to be little good evidence about disparities in the VA. I might be clearer about the limitations of the VA data on disparities as you discuss an intervention to improve them in KQ3.	Noted

Comment	Response
Note also that in the first sentence the word “joint” probably doesn’t belong there or you need to add TJR	Sentence has been verified.
Page 26 paragraph 4----since you are talking about the disparities, even if the data are not robust, I think you might as well say what you found in terms of the disparities. Also, I thought some of the differences were decreased with adjustment for confounders? This is a good place to reiterate that.	This is discussed in the “clinical need” paragraph. The differences may decrease after adjustment, but they don’t go away entirely.
Page 27----KQ3. If you are going to talk about the one published study it makes sense to me to summarize what it showed	This has been added.
Recommendations for Future Research	
It would be appropriate to have more specific recommendations at the end of the review, arising from the evidence, or lack thereof. For instance: 1) areas with conflicting findings; 2) areas needed to be studied in Veterans, for which little information is available (e.g. women are mentioned, how about Hispanics); and 3) potential interventions that should be evaluated on the basis of the findings – patient-based or QI.	We have revised the future research section
This is a very comprehensive and detailed review of the literature, and it would be very helpful if more specific recommendations could be drafted in summary of the review. There may also be recommendations that could be made with respect to implementation, but given the current state of the research, it seems that more evidence base is needed regarding interventions to address disparities, before these are put into clinical practice. I think it will help readers / stakeholders to get more out of the evidence synthesis if a more detailed “take home” message is provided with respect to what is still needed.	We have revised the future research section
The report does not identify anything to implement. The call for more research seems appropriate	We have revised the future research section
I am not sure I agree with the recommendations for further research. it seems to me that if the evidence base is limited for first and second generation disparity studies that these should be conducted prior to suggesting more third generation research. it is not completely clear to me that there are disparities at the VA. I would like to see a Discussion section (it can be short) in this paper with some discussion of the problems with this evidence base. In particular, I am struck by how often point estimates of disparity were either reduced or eliminated by adjusting for confounders. I think this deserves more synthesis and discussion.	We have revised the future research section
Appendix F	
Appendix F. Number of articles is 57? The number does not match up with the numbers in Figure 3	Numbers have been updated.

APPENDIX F. EVIDENCE TABLES

Key Question 1/ Generation 1 Evidence Table

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender, Race**/ How determined	Joint Discussed***	Outcomes	Results
VA Studies								
Hausmann; 2010 ²⁶	2005-2008	Cohort/ Case-control; Unclear; Not specified	457; 457	VA, multi-institution	W/ Self-Reported	TKR, THR	recommendation for procedure; receipt of procedure	Lower odds of receiving a TJR recommendation for B than W of similar age and disease severity (OR 0.46, [95% CI 0.26–0.83]; P=0.01). Difference was not significant adjusting for patient preference for TJR (OR 0.69, [95% CI 0.36–1.31], P=0.25). TJR less likely for B than W of similar age and disease severity (OR 0.41 [95% CI 0.16–1.0], P=0.06); difference reduced adjusting for recommendation for procedure at the index visit (OR 0.57 [95% CI 0.21–1.54], P=0.27).
Borrero; 2006 ⁴⁵	1999	Cross-sectional; Pop based; N/A	329,461; 329,461	VA National	Women/ Admin	TKR, THR	Adjusted odds of getting TJR	Among patients with OA, men and women in the VA were equally likely to undergo TKR (153 [1.6%] women and 4,638 [1.5%]) men and THR (73 [.8%] women and 2147 [.7%] men). Receipt of surgery within 2 years for women with OA versus men was not significant (TKR: OR 0.96 [95% CI 0.82 to 1.13]) and (THR: OR 0.99 [95% CI 0.79 to 1.26]).
Jones; 2005 ²⁷	1999	Cohort/ Case-control; Pop based; Not specified	260856; 260856	VA National	W, B/ Admin. data	TKR	receipt of procedure	B were less likely than W to have received TKA within 2 years (OR 0.72, [95% CI 0.65–0.80] in OA cohort and OR 0.72, [95%CI 0.63–0.81] in specialty clinic subcohort.
Non-VA Studies								
Hawkins; 2011 ⁵⁴	2006-2007	Cross-sectional; Pop-based; N/A	2.9 million; 0	Medigap	% non-white by zip code/ Admin	Hip or knee replacement	receipt of procedure	Patients living in high-minority areas were 20% less likely to undergo a hip or knee replacement as low minority areas.
Bang; 2010 ⁵⁶	1996-2005	Cross-sectional; Pop-based; N/A	8000000	NIS	W, B, H, A/ Admin	TKR, THR	receipt of procedure	Non-whites had lower odds of THA and TKA compared with whites. Minorities were 23% to 64% less likely to undergo arthroplasties. Racial disparities were larger than income disparities and not confined to elderly or low-income.

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender, Race**/ How determined	Joint Discussed***	Outcomes	Results
Francis; 2009 ⁶¹	2005	Cross-sectional; Pop-based; N/A	46000000; 0	Medicare, NIS	W, B, H, A/ Admin. data	TKR, THR	receipt of procedure	Compared with urban beneficiaries, rural were more likely to have TJR (OR 1.27 [95% CI 1.26– 1.28]). Adjusting for age, sex, race/ethnicity, income, poverty ratio, and state, rural beneficiaries were still 14% more likely to have TJR (OR 1.14 [95% CI 1.13–1.16]).
2009 ⁵¹	2000/2006	Cross-sectional; Pop-based; N/A	26000000; N/A	Medicare	W, B/ Admin. data	TKR	receipt of procedure	From 2000 to 2006, TKR rate in the US increased 58%, from 5.5 to 8.7 per 1,000, with similar increases among W (61%) and B (56%). Rate of TKR for B was 37% lower than W in 2000 (3.6 versus 5.7 per 1,000) and 39% lower in 2006 (5.6 versus 9.2 per 1,000).
Basu; 2008 ⁵⁹	1997-2001	Cross-sectional; Pop-based; N/A	71418; NR	National, HCUP	W, B, H/ Admin. data	THR	receipt of procedure	No difference in the likelihood of THR between B, W and H for 1997 or 2000, after adjusting for income, urban/rural, distance from hospital, and social isolation, but not severity of arthritis.
Steel; 2008 ⁶²	1998, 2000, 2002	Cohort/ Case-control; Pop-based; Not specified	14807; NR	Health and retirement study, national	W, B/ Self-reported	TKR, THR	need for surgery; receipt of procedure	Lower receipt of TJR in B (vs W: OR 0.47; CI 0.26– 0.83) or less educated (0.65; 0.44–0.96). Differences not explained by employment, access, family responsibilities, disability, living alone, comorbidity, or excluding younger than Medicare.
Hanchate; 2008 ⁵⁷	1994-2004	Cohort/ Case-control; Pop-based; Not specified	18439; NR	National	W, B, H/ Self-reported	TKR	receipt of procedure	B men (relative to W women) were less likely (OR 0.46 (0.28–0.78), [$P < 0.05$]) to receive TKA. Adjusting for economic factors, racial/ethnicity, TKA rates differences for women disappeared, while remaining large for B men (OR 0.56 [0.33–0.95]).
Skinner; 2006 ⁵³	2000 (Medicare)	Cross-sectional; Pop-based; N/A	27494659; NR	NHANES, national Medicare	W, B, H, A/ Admin. data	TKR	receipt of procedure; Prevalence of OA	Relative to W men, B men were less likely to undergo TKA (OR 0.36 [95% CI 0.34 to 0.38); as were H men (OR 0.67 [0.62 to 0.73]; Asian men (OR 0.28 [0.24 to 0.32]; and Asian women (OR 0.45 [0.41 to 0.49]. W women were more likely (OR 1.34 [1.33 to 1.36]). [No income gradient for clinical and radiographic measures of arthritis, except a negative association of income and pain on passive motion ($P < .05$).]

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender, Race**/ How determined	Joint Discussed***	Outcomes	Results
Jain; 2005 ⁵⁵	1990-2000	Cross-sectional; Pop-based; N/A	443008; NR	NIS, national	W, B, H/ Admin. data	TKR	receipt of procedure	In 1998-2000 as compared to 1990-1993, B and Hispanic patients were more likely to undergo TKA (OR 1.6 [95% CI 1.5–1.6] and OR 2.7, [95% CI 2.5–2.9], respectively). However, W patients accounted for 87.5% and 93.0% of TKAs, in the 2 time periods.
Jha; 2005 ⁵²	1992-2001	Cross-sectional; Pop-based; N/A	29000000; NR	Medicare	Non-B/ Admin. data	THR	receipt of procedure	Rates of TKR and THR among the Medicare fee for service population were compared from 1992 to 2001. Women had higher age-adjusted rates of procedure use than men, and nonblacks had higher rates than blacks. In 2001, nonblack men had a rate of 5.05/1000 population for TKR, compared to 1.85 for black men. Among women, rates of TKR per 1000 population were 6.6 among nonblacks and 5.1 among blacks.
Mehrotra; 2005 ¹⁶	1990-2000	Cross-sectional; Pop-based; N/A	67,475; NR	Regional (Wisconsin Hospital Discharges)	Gender/ Admin. data	TKR	receipt of procedure	In both 1990 and 2000, women had higher rates of TKR. Rates of TKR per 100,000 in 1990 were 30 for women compared to 23 in men, and in 2000 were 46 in women compared to 35 in men.
Olson; 2005 ⁵⁸	1993-2001	Cross-sectional; Pop-based; N/A	Many; NR	Regional	W, B, H/ Admin. data	TKR	receipt of procedure	Connecticut hospital data (1996-1998) found that age adjusted rates per 100K discharges for TKR was highest for black women (115.8, 95% CI 103.9-127.7) and lowest for black men (44, 34.9-68.9) and Hispanic men (16.9, 10.1-23.8) and women (47.5, 37.8-57.2). White women had rates of 84.9 (82.4-87.4) and men 66.5 (63.9 -68.9).
Skinner; 2003 ⁴⁷	1998-2000	Cross-sectional; Pop-based; N/A	403251; 0	National, Medicare	W, B, H/ Admin. data	TKR	receipt of procedure	Rate of TKA was higher for W women (5.97 procedures per 1000) than for H women (5.37 per 1000) and B women (4.84 per 1000). Rate for W men (4.82 procedures per 1000) was higher than H men (3.46 per 1000) and more than double that for B men (1.84 per 1000). The rates were lower for B men in nearly every region of the country ($P<0.05$). [For H population and for B women, racial/ethnic disparities were due in part to geographic differences rather than to differences in the rates for racial and ethnic groups within geographic areas. Residential segregation and low income levels contributed to disparities.]

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender, Race**/ How determined	Joint Discussed***	Outcomes	Results
Escalante; 2002 ⁴⁸	95-96	Cross-sectional; Pop-based; N/A	19311; 0	Medicare in NM, IL, TX, AZ	W, B, H, A/ Two or more sources	THR	receipt of procedure	1% THR recipients and 3.3% controls were H ($P < .001$). Odds of THR decreased as probability of H ethnicity increased (OR 1.00 non-H surnames to OR 0.36 H surnames (95% CI, 0.31, 0.43). Poverty did not modify the low odds of THR among H (OR, 0.25 Medicaid-eligible Hispanic persons; 95% CI, 0.19, 0.33; and OR, 0.30 Hispanic persons not Medicaid eligible; 95% CI, 0.24, 0.38).
Oishi; 1998 ⁶⁰	85-89	Cross-sectional; NR; N/A	754; 0	Regional	W, A/ Medical record review	THR	receipt of procedure	THR for W was three to 25 times greater than that of Japanese, Chinese, Filipino, and Hawaiians. Risk of THR for W women was 4.4%, compared with 1.1% for Japanese women and 1.7% for Chinese women. For white men, the incidence rate is 3.6%, which is 4.5 to nine times greater than the rate for other ethnic groups. Some between region differences were noted (Hawaii versus San Francisco).
Giacomini; 1996 ¹⁴	1989-1990	Cross-sectional; Pop-based; N/A	6586; NR	OSHPD, regional	W, B, H, A/ NR	THR	receipt of procedure	Asians had higher odds of THR (OR 2.13 [95% CI 1.3-3.45]) than W. W had higher, but non-significant, odds of THR than H (OR 1.32 [.87-1.96]) and than B (OR 1.56 [.97-2.50]).
Katz; 1996 ⁴⁹	85-90	Cross-sectional; Pop-based; N/A	414079; 0	Medicare, national	W, B/ Two or more sources	TKR	receipt of procedure	Odds of W receiving TKR were 1.5 times greater than for B. Adjusting for demographic factors, regional variation remained. TKR were over two and one-half times more likely for B women than for men (OR 1.66); the difference was only 24 percent for W women versus W men (OR = 1.24). Procedures were performed on W men much more often than on B, (OR 2.50). Difference between W and B women was much smaller (OR = 1.16).
Hoaglund; 1995 ¹²	84-88	Cross-sectional; Pop-based; N/A	1589; 0	San Francisco	W, B, H, A/ Medical record review	THR	receipt of procedure	The greatest annual rate of THR occurred in W women (97 per 100 000), followed by W men, B women, B men, H women, and H men. Smallest numbers were found in Asians, rate was 10% of W. Age standardized THR rates for primary coxarthrosis per 100 000 were greatest among W (43.0) and least among Asians (1.3 for Chinese). Mean age undergoing THR for primary coxarthrosis was 70 years for W and a decade younger in other groups.

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender, Race**/ How determined	Joint Discussed***	Outcomes	Results
McBean; 1994 ⁵⁰	86-92	Cross-sectional; Pop-based; N/A	a lot; 0	Medicare, national	W, B/ Admin. data	TKR, THR	receipt of procedure	TKR increased from 1986 and 1992, 98% among white beneficiaries and 121 percent among blacks. In 1992, the rate in blacks was 64% as great as for whites.
Wilson; 1994 ¹¹	1980-1988	Cross-sectional; Pop-based; N/A	over 3000; NR	NHANES, Medicare	W, B/ Admin. data	TKR	receipt of procedure; Rate of OA	B were less often treated with TKR than W (men: OR=3.16 [1.69-5.91]; women: OR=1.55 [1.00-2.41]) for age 65-69.
Escarce; 1993 ⁸	1986	Cross-sectional; Pop based; N/A	1204022; 0	Medicare, national	W, B/ Admin. data	TKR, THR	receipt of procedure	W are two-fold more likely to undergo THR (RR 2.36 [1.92, 2.89]) or TKR (RR 2.02 [1.63, 2.49]) than blacks.

Key Question 2/ Generation 2 Evidence Table

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender/ Race**/ How determined	Joint Discussed***	Outcomes	Results
Direct Comparison of VA and Non-VA Studies								
Ang; 2009 ³⁰	2003-2006	Cohort/ Case-control; Convenience; 91.4%	676; 388	Single VA	W, B/ Self-reported	TKR, THR	Appropriateness, benefits, barriers, OA severity, length of time to referral; outcome expectations of surgery, perceived risk	Clinical appropriateness (HR 1.95, [95% CI] 1.15– 3.32; $P < 0.01$) predicted referral to orthopedic surgery. Neither race (HR 1.30, 95% CI 0.94–2.05; $P = 0.1$) nor health beliefs (HR 1.0, $P = 0.5$) were associated with referral status.
VA Studies								
Hausmann 2011 ²⁹	2005-2008	Cohort/ Case-control; Unclear; Not Specified	409; 409	VA	W, B/Self-reported	TKR, THR	Patient-provider communication	Visits with B, compared with W, contained less discussion of biomedical topics ($B = -9.14$, 95% CI -16.73 - -1.54) and more rapport-building statements ($B = 7.84$; 95% CI 1.85- 13.82. No racial differences in length of visit, overall amount of dialogue, patient activation/engagement statements, discussions of psychosocial issues, physician verbal dominance, displays of positive affect, or evidence of informed decision making.

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender/ Race**/ How determined	Joint Discussed***	Outcomes	Results
Hausmann; 2010 ²⁶	2005-2008	Cohort/ Case-control; Unclear; Not specified	457; 457	VA	W, B/ Self-reported	TKR, THR	recommendation for procedure; receipt of procedure	Lower odds of receiving a TJR recommendation for B than W of similar age and disease severity (OR 0.46, [95% CI 0.26–0.83]; P=0.01). Difference was not significant adjusting for patient preference for TJR (OR 0.69, [95% CI 0.36–1.31], P=0.25). TJR less likely for B than W of similar age and disease severity (OR 0.41 [95% CI 0.16–1.0], P=0.06); difference reduced adjusting for recommendation for procedure at the index visit (OR 0.57 [95% CI 0.21–1.54], P=0.27).
Jones; 2008 ³³	999	Cross-sectional; Convenience; N/A	939; 939	VA	W, B/ Self-reported	TKR, THR	prayer for pain, coping strategies, self efficacy	B more likely to perceive prayer helpful (OR 3.38, 95% CI [2.35 to 4.86]) and use prayer (OR 2.28, 95% [1.66 to 3.13]) to treat osteoarthritis pain as compared to W. B more likely to use coping and praying ($\beta=0.74$, 95% CI [0.50 to 0.99]).
Groeneveld; 2008 ³⁴	2004-2006	Cross-sectional; Convenience; N/A	909; 909	VA	W, B/ Self-reported	TKR, THR	outcome expectations of surgery	B knee OA patients have lower expectation score (scale 0-76) than W even with adjustment for disease severity, SES, social support, literacy and trust (difference -3.8 points [95% CI 1.2, 6.3], and 4.2 points (95% CI 0.4, 8.0) among hip patients.
Weng; 2007 ⁴⁴	999	Experimental; Convenience; Not specified	64; 64	VA	W, B/ Self-reported	TKR	outcome expectations of surgery, willingness to consider surgery, knowledge of surgery, alternative treatment	B had lower (but not significant) expectations for TKR than W for pain (WOMAC score 41 versus 34; P=0.18) and physical function (WOMAC score 38 versus 30; P = 0.13). B were less likely to have heard of TKR (49% versus 72%; P=0.02) and less likely to know someone who had TKR (34% versus 53%; P = 0.05) than W.
Ibrahim; 2005 ²⁸	1996-2000	Cohort/ Case-control; Pop-based; N/A	18811; 18811	NSQIP, VA	W, B, H/ Admin. data	TKR, THR	complications	Rates of non-infection and infection-related complications after TKA were higher among B compared with W (RR 1.50, [95% CI 1.08–2.10] and RR 1.42, [95% CI 1.06–1.90]). H had a higher risk of infection-related complications (RR 1.64, 95% CI 1.08–2.49) relative to W. Race/ethnicity was not associated with the risk of non-infection-related or infection-related complications for THR. 30-day mortality was 0.6% following TKA and 0.7% following THR, with no race/ ethnicity differences

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender/ Race**/ How determined	Joint Discussed***	Outcomes	Results
Lopez; 2005 ³⁶	1997-2000	Cross-sectional; Convenience; 728/770	596; 596	VA	W, B/ Self-reported	TKR, THR	participant/provider relationship, perception access to care, receipt of referral	B were less likely than W to report difficulty getting medical care (OR 0.54 [0.34-0.88]). B were less likely than W to perceive the patient- physician relationship as excellent (24.7% vs. 36.3%, $P<0.01$) and less likely to have confidence in their primary physician (75.7% vs. 82.6%, $P=0.04$). Difficulty accessing care outside VA was not different between groups (52.4% vs. 52.2%, $P=0.95$).
Ang; 2003 ³⁷	999	Cross-sectional; Convenience; 38 lost to follow up	558; 558	VA	W, B/ Self-reported	TKR, THR	perception of symptoms	B and W were not different in mean scores for WOMAC pain and WOMAC function when stratified by joint space narrowing, osteophyte and Kellgren Lawrence grades. After controlling for important covariates, ethnicity was not a significant predictor of WOMAC pain and function.
Ibrahim; 2003 ³⁸	999	Cross-sectional; Convenience; Not specified	300; 300	VA	W, B/ Self-reported	TKR, THR	Perception of pain	B and W patients describe the quality of their chronic knee and hip pain differently. Chronic pain quality descriptions correlate with western Ontario and McMaster Universities Arthritis Index Scores but not radiologic stage of disease. {factor analyses}
Ibrahim; 2002 ³⁹	1997-2000	Cross-sectional; Convenience; 738/776	596; 596	VA	W, B/ Self-reported	TKR, THR	willingness to consider surgery, outcome expecta- tions of surgery, familiarity with surgery	B were less likely than W to be willing to consider surgery for severe arthritis (OR 0.53, [95% CI 0.30-0.96]). After adjustment for outcome expectations, the difference between races in willingness to consider was not significant (OR 0.86, [95% CI 0.45-1.63]).
Ang; 2002 ⁴⁰	97-00	Cross-sectional; Convenience; 95%	596; 596	VA	W, B/ Self-reported	TKR, THR	Role of prayer in the management of arthritis, willingness to consider surgery	B less willing than W to consider surgery for severe hip or knee arthritis pain (OR .059, [95% CI 0.34-0.99]). B more likely than W to perceive prayer as helpful in managing their arthritis (OR 2.1; [95% CI, 1.19, 3.72]).
Ibrahim; 2002 ⁴¹	97-00	Cross-sectional; Convenience; 738/776	596; 596	VA	W, B/ Self-reported	TKR, THR	QOL	For patients with chronic joint disease, B less likely than W to rate quality of life as excellent or very good. Difference persisted after adjusting for demographic, clinical, and psychosocial covariates, and severity of osteoarthritis (B=-0.121, $P=.004$).

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender/ Race**/ How determined	Joint Discussed***	Outcomes	Results
Ibrahim; 2002 ⁴²	97-00	Cross-sectional; Convenience; 738/776	596; 596	VA	W, B/ Self-reported	TKR, THR	outcome expectations of surgery, knowledge of joint replacement	B were less likely than W to have family/friends that had TJR (OR 0.39 [.26-.61]) or a good understanding of TRJ (OR 0.62 [.42-.92]). B more likely to expect longer hospital course (OR 4.09 [2.57-6.54]), moderate to extreme pain (OR 2.61 [1.74-3.89]), moderate to extreme difficulty walking after joint replacement (OR 2.76 [1.83-4.16]).
Ibrahim; 2001 ⁴³	97-00	Cross-sectional; Convenience; 738/776	593; 593	VA	W, B/ Self-reported	joint replacement	efficacy arthritis treatment	B were more likely than W to perceive TJR as efficacious (OR .52 [.28-.98]) and more likely to rely on self-care measures for their arthritis (OTC meds: OR 1.76 [1.14-2.72]); friend/family advice: OR 2.11 [1.44-3.07]); decrease activities: OR 2.22 [1.28-3.85]); apply med cream: OR 2.27 (1.38-3.73]). Use of prayer more likely to be perceived as efficacious in B (OR 1.93 [1.19-3.14]).
Non-VA Studies								
Kamath; 2010 ⁷⁷	2004	Cohort/ Case-control; Consecutive patients; Not specified	185; N/A	Single institution	B, Non-B/ Medical Record Review	TKR	outcome of procedure, outcome expectations of surgery	B men had longer delays to presentation than non B men (29.9 months [CI 17.2, 42.6] vs 20.0 months [CI 4.4, 35.6]) and worse 2-year KSS (89.6 months [CI 85.0, 94.2] vs 94.1 months [CI 91.2, 97.0]). B women had worse final ROM and similar final gains in ROM (postoperative minus preoperative) controlling for confounders.
Slover; 2010 ⁷⁸	1997-2006	Cross-sectional; Consecutive patients; Not specified	3542; 0	Single institution	W, B, H/ Self-reported	TKR, THR	preop jt function	Lower function with Harris Hip Scores 4.9 (P<.0001) and 8.77 (P<.001) and Knee Society Scores that were 6.03 (P<.06) and 12.8 (P<.001) points lower in B and H patients than W.
Suarez-Almazor; 2010 ⁷³	999	Qualitative; Unclear; Not specified	37, 0	Single institution	W, B, H/ Self-reported	TKR	willingness to consider surgery, outcome expectations of surgery, TKR Knowledge, Current prob knee OA	Attitudes and beliefs of surgical decision-making were primarily based on personal experiences. Personal experiences had both positive or negative impacts and included concerns about outcomes following surgery and possible complications. B did not have more concerns or fewer expectations.

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender/ Race**/ How determined	Joint Discussed***	Outcomes	Results
Epstein; 2010 ⁸¹	2001-2004	Cross-sectional; Pop-based; Not specified	25598, 0	NYC discharge data	W, B, H, A/ Admin. data	THR	use HVH (hospital), use HVH (surgeon) receipt of procedure	B, Asians, H were more likely to be operated for TJR in low volume hospitals by low volume surgeons than whites (25.3%, 35.0%, 23.0% and 15.6%; P<.001)
Lavernia; 2010 ⁷⁴	2000-2002	Cohort/ Case-control; Convenience; Not specified	331; 0	Single institution	W, B/ Self-reported	TKR, THR	outcome of procedure, fear of surgery, physical function	B patients had greater fear before joint arthroplasty compared with W. After surgery, B had higher fear subscale, cognitive subscale, and total PASS score (WOMAC physical function, pain, and total scores.
Berges; 2008 ⁸⁰	2002-2003	Cohort/ Case-control; Pop-based; N/A	69793; NR	National	W, B, H, A/ Admin. data	THR	outcome of procedure	B and H had higher odds of discharge to home following hip replacement (B: OR 1.23 [1.107-1.41] and H: OR 1.5 [1.15-1.99]). B not significant. Men had higher odds of discharge to home (OR 1.18 [1.01-1.17]). Mean functional status change not predictive of discharge disposition (OR 1.10 [1.10-1.11])
Steel; 2008 ⁶²	1998, 2000, 2002	Cohort/ Case-control; Pop-based; Not specified	14807; NR	National	W, B/ Self-reported	TKR, THR	need for surgery, receipt of procedure	Lower receipt of TJR in B (vs W: OR 0.47; CI 0.26–0.83) or less educated (0.65; 0.44–0.96). Differences not explained by employment, access, family responsibilities, disability, living alone, comorbidity, or excluding younger than Medicare.
Hanchate; 2008 ⁵⁷	1994-2004	Cohort/ Case-control; Pop-based; Not specified	18439; NR	National	W, B, H/ Self-reported	TKR	receipt of procedure	B men (relative to W women) were less likely (OR 0.46 (0.28–0.78), [P< 0.05]) to receive TKA. Adjusting for economic factors, racial/ethnicity, TKA rates differences for women disappeared, while remaining large for B men (OR 0.56 [0.33–0.95]).
Kroll; 2007 ⁷⁶	999	Qualitative; Convenience; Not specified	37; NR	Single institution	W, B, H/ Self-reported	TKR	attitudes and beliefs about TKA	Knee OA is experienced differently by ethnicity and groups, and perceptions of the cause of knee OA vary. Trust is important for H considering TKA. Economic factors do not constrain the decision to have surgery.

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender/ Race**/ How determined	Joint Discussed***	Outcomes	Results
SooHoo; 2011 ⁸³	1995-2005	Cross-sectional; Pop-based; N/A	138399; NR	OSHDD (California input database)	W, B, H, A/ Admin. data	THR	receipt by hospital volume	H had higher RRR [3.52 (95% [CI] 2.61-4.74; $p < .001$) for the use of a low-volume hospital when compared to W. B (RRR, 1.78; $p = .023$) and Asian patients (RRR, 1.77; $p = .048$) also had a higher RRR compared to W for the use of low-volume hospitals
Liu; 2006 ⁸²	2000-2004	Cross-sectional; Pop-based; N/A	1E+05; NR	OSHDD (California input database)	W, B, H, A/ Admin. data	TKR	receipt by hospital volume	B, Asians, and H were more likely to have TKR at low volume hospital than W (RR=1.32 [95% CI 1.25-1.39], RR=1.72 [95% CI 1.60-1.81], RR=1.64 [95% CI 1.58-1.69])
Byrne; 2006 ⁸⁴	999	Cross-sectional; Random digit dialing; Not specified	391; NR	Single institution	W, B, H/ Self-reported	TKR	willingness to consider surgery	B less likely to chose surgery than W (OR 0.63 [CI 0.42, 0.93]). Women and older patients were also less likely to choose surgery (OR 0.69 [0.51, 0.94], OR 0.98 [0.97, 0.99]). Larger reductions in negative symptoms with surgery increased the likelihood of choosing surgery. No difference between the public and patients, and no effect of income level was noted.
Suarez-Almazor; 2005 ⁶⁹	2001-2002	Cross-sectional; Pop-based; N/A	198; 0	Single institution	W, B, H/ Self-reported	TKR	recommendation for procedure, willingness to consider surgery, outcome expectations of surgery, preferences for surgery, familiarity of surgery	Physician more likely to discuss TKR with B (27%), 15% W, 11% H ($P=.04$). More W than minorities (B and H combined) considered TKR (42% vs 28%; $P=.04$). No differences between B, H, W being familiar with TKR.
Figaro; 2005 ⁷⁰	999	Cross-sectional; Convenience; 104/114	94; 0	Harlem	B/ Self-reported	TKR	outcome expectations of surgery	In B with high rate of severe OA (mean QoL 7.6 ± 1.7), few (36%) believed TKR would improve knee pain; and 45% felt surgery would not improve their health.
Byrne; 2004 ⁸⁵	2001	Cross-sectional; Pop-based; 23%	193; 0	Harris County	W, B, H/ Self-reported	TKR	willingness to pay	Willingness to pay (WTP) as a percentage of income was lowest for B (16.7% for mild OA) as compared to 32.9% W, 26.4% H. Controlling for income, differences in WTP between B and W were significant in multivariate regression analyses, whereas values for H and W were not.

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender/ Race**/ How determined	Joint Discussed***	Outcomes	Results
Figaro; 2004 ⁷¹	999	Qualitative; Convenience; Not specified	94; 0	Harlem	B/ Self-reported	TKR	outcome expectations of surgery, briefs and goals surgery, pref stay in current state, relationship w specialist	Content analyses identified 6 themes: preference for natural remedies, negative expectations of surgery, beliefs about God's control, preference for continuing in their current state, relationships with specialists, and fear of surgery or death.
Chang; 2004 ⁷²	1998-1999	Qualitative; Consecutive; N/A	37; 0	Single institution	W, B/ Self-reported	TKR	concern about surgery	B women asked the most questions about criteria for TKR; W women asked about drawbacks from surgery; W men asked about devices; B men asked about financial issues and insurance coverage. Only W asked about intraoperative issues. W women asked about recuperation, functional recovery and pain, B women asked about long-term outcomes and support after surgery. W men asked about QOL and B men asked no questions. W men had greatest factual knowledge about surgery.
Ottenbacher; 2003 ⁷⁹	1994-1998	Cross-sectional; Pop-based; N/A	12328; NR	4DSMR, national	W, B, H, A/ Two or more sources	TKR, THR	outcome of procedure	W and B were ($P <0.05$) more likely to be discharged home alone and responsible for their own care than Asian or H. 36% H after THA or TKA received inpatient medical rehabilitation 58% W, 67% B, and 56% Asians.
Blake; 2002 ⁷⁵	999	Cross-sectional; Pop-based; 44%	970; 0	Medicare Manhattan	W, B/ Two or more sources	hip/knee surgery	Social network, Perception of benefit of arthritis treatment	42% B compared 65% W reported knowing someone who had surgery for hip or knee pain ($P <.0001$). B less likely than W to report that surgery had helped someone they knew with hip or knee pain (but not significant). B more likely to have sought care in ER/clinic 22% vs 9%, $P <.005$ and less likely to have seen an orthopedic surgeon 3% vs 15%, $P <.0001$). No racial differences in use of self-treatments (OTC, herbs PT, health/cold)

Author; Date	Data Dates*	Study Type; Sample Selection; Response Rate/ Follow Up	Total pop; VA pop*	Data Source	Gender/ Race**/ How determined	Joint Discussed***	Outcomes	Results
Wilson; 1994 ¹¹	1980-1988	Cross-sectional; Pop-based; N/A	over 3000; NR	NHANES, Medicare	W, B/ Admin. data	TKR	receipt of procedure, Rate of OA	Prevalence of symptomatic OA knee was lower (but not significant) in W compared to B (men OR .39 [.13-1.14] and women OR .78 [.34-1.80]). Racial differences in TKR were consistent across income levels and were unexplained by B having operations at an earlier age or using competing procedures.

Key Question 3/ Generation 3 Evidence Table

Author; Date	Data Dates*	Study Type	Total pop; VA pop*	Data Source	Race**/ How determined	Joint Discussed***	Outcomes	Results
Weng; 2007 ⁴⁴	999	Experimental; Convenience; Not specified	64; 64	VA	W,B/ Self-reported	TKR	Willingness to consider surgery, outcome expectations of surgery willingness to consider surgery, knowledge of surgery, alternative treatment	At baseline, 13% W and 29% B were willing to consider surgery ($P < 0.12$); after intervention, 13% W and 33% B were willing to consider surgery ($P < 0.06$).

*No data denoted as NR or 999

** W= White; B= Black; H= Hispanic; A=Asian

*** TKR=Total Knee Replacement; THR=Total Hip Replacement