

This manuscript applies the GRADE Evidence-to-Decision (EtD) framework to develop clinical practice recommendations for the surveillance of non-malignant asbestos-related diseases (NMARDs) in asbestos-exposed populations within Colombia. The topic is of clear public health importance, and the use of a structured EtD approach is methodologically appropriate. However, the current version suffers from critical weaknesses in evidence base, methodological reporting, and the alignment between recommendation strength and evidence certainty. These limitations undermine the reliability and implementability of the proposed guidelines. My comments are as follows:

1. The search strategy for Questions 2 and 3 is poorly designed, yielding an unacceptably thin evidence base. The use of MeSH/free-text terms such as "Watchful Waiting," "Critical Pathways," and "Eligibility Determination" is poorly aligned with the clinical context of asbestos post-screening follow-up. This resulted in zero retrievals from Embase and the Cochrane Library, leaving only two and three included studies for Questions 2 and 3, respectively. Such a sparse evidence base cannot support definitive recommendations regarding follow-up intervals (e.g., "every three years") or the management of positive findings.

It's suggest to redesign the search strategies for Q2 and Q3 using relevant terms (e.g., follow-up, surveillance, monitoring, screening interval, radiographic progression, HRCT). If the evidence base remains insufficient after re-searching, the specific follow-up interval must be explicitly labeled as a good practice statement based on expert consensus, not as an evidence-driven recommendation. The rationale for the three-year interval (e.g., latency periods, resource considerations) should be transparently stated.

Response 1

Questions 2 and 3, the definition of the included MeSH terms was:

Watchful Waiting: *Scope Note* - Clinical management approach wherein immediate therapy is not provided but there is a period of observation during which periodic tests monitor patient and the progression of the illness. (Driffield T, Smith PC Med Decis Making. 2007 Mar-Apr;27(2):178-88) (Entry Term(s): Active Surveillance, Expectant Management. Previous Indexing Follow-Up Studies) (<https://meshb.nlm.nih.gov/record/ui?ui=D057832>)

Critical Pathways: *Scope Note* - Schedules of medical and nursing procedures, including diagnostic tests, medications, and consultations designed to effect an efficient, coordinated program of treatment. (From Mosby's Medical, Nursing & Allied Health Dictionary, 4th ed) (<https://meshb.nlm.nih.gov/record/ui?ui=D019091>)

Eligibility Determination: *Scope Note* - Criteria to determine eligibility of patients for medical care programs and services. (<https://meshb.nlm.nih.gov/record/ui?ui=D004608>)

Terms such as follow-up, surveillance, and monitoring were included within the MeSH term "watchful waiting". For searches conducted in questions 2 and 3, we used a limited number of terms but we believe those are related; additionally, we ran them without time restrictions or additional filters in order to develop a broad search strategy.

We ran the search again with the suggested free terms, leaving negative and positive findings as a new line of search, and we could not identify relevant studies in Pubmed. The subject matter experts on the working group believe that the limited evidence identified through this search is due to a lack of studies addressing these issues.

History and Search Details

 Download
  Delete

Search	Actions	Details	Query	Results	Time
#6	...	>	Search: #1 AND #2 AND #5	6	05:16:56
#5	...	>	Search: Positive findings [Title/Abstract]	8,996	05:05:34
#4	...	>	Search: #1 AND #2 AND #3	2	05:01:12
#3	...	>	Search: Negative findings [Title/Abstract]	4,261	05:00:55
#2	...	>	Search: ((((((((((Watchful waiting[Title/Abstract]) OR (Watchful waiting[MeSH Terms])) OR (Critical Pathways[Title/Abstract])) OR (Critical Pathways[MeSH Terms])) OR (Eligibility Determination[Title/Abstract])) OR (Eligibility Determination[MeSH Terms])) OR (follow-up[Title/Abstract])) OR (surveillance[Title/Abstract])) OR (monitoring[Title/Abstract])) OR (screening interval[Title/Abstract])) OR (radiographic progression[Title/Abstract])) OR (HRCT[Title/Abstract])	2,572,519	05:00:40
#1	...	>	Search: ((Asbestos[Title/Abstract]) OR (Asbestos[MeSH Terms])) OR (Exposure to asbestos[Title/Abstract])	16,436	04:59:35

Showing 1 to 6 of 6 entries

To explicitly address the issue of the three year interval, in Line 375, the wording was modified to: "...three-year interval having been defined by consensus within the working group".

2.The strength of recommendations is inconsistent with the judged certainty of evidence. The working group formally judges the certainty of evidence as Low and resource use as "Don't know" (Section 3.7). Nevertheless, the Conclusions state that "The working group recommends CXR with ILO classification as the primary screening tool" and advises routine three-year follow-up. Under GRADE principles, low-certainty evidence should yield conditional recommendations rather than strong ones.

It's suggest to revise all recommendation statements so that their strength accurately reflects the underlying evidence certainty. Use "We conditionally suggest..." or "The working group suggests..." for recommendations supported by low-certainty evidence. Add a dedicated paragraph in the Discussion explaining why the balance of effects favors the intervention despite low certainty.

Response 2

Thank you for noticing; in Line 347, the word "conditional" was added, and 'development team' was changed to 'working group' to align the term with the rest of the document. In Line 457 the word "clear recommendation" was changed to "conditional recommendation". In Line 462 the word "conditionally" was added.

3.The terms "screening" and "surveillance" are used interchangeably throughout the manuscript, creating conceptual confusion. The authors acknowledge in Section 4.3 that current evidence is insufficient to support formal population screening according to Wilson & Jungner principles. Yet the title, abstract, PICO questions, and main text repeatedly use "screening" (e.g., "optimal screening test," "Mass Screening" as a MeSH term). This contradiction weakens the manuscript's conceptual clarity.

Recommendation: Standardize terminology throughout. Use "initial surveillance evaluation" or "baseline assessment" for the first encounter and "periodic surveillance" or "follow-up monitoring" for subsequent evaluations. If the term "screening" is retained, its specific meaning within Colombia's legally mandated care pathway (RIAS) must be explicitly defined in the Introduction to distinguish it from classical population screening.

Response 3

Through previous modifications to the manuscript we had intended that this issue had been appropriately and explicitly addressed but we agree that there were still several instances where the wording was insufficiently clear. As a result, we have reviewed all instances where the word 'screening' was used. As a result, we have made several changes:

At Line 29, the text has been amended to read: "It establishes evidence-informed surveillance (screening and follow-up protocols) for..."

At Lines 104-106: the following was added: "This resolution defined screening as an initial intervention within the RIAS that should be covered and paid for by the healthcare system."

The last paragraph before the Methods section now reads "We use the term "screening" to refer to the process of attempting to detect asbestos-related disease within the asbestos-exposed population rather than implying the use of a comprehensive program of early identification, investigation and management of asbestos-related disease."

In other places where the word 'screening' is used, we believe it is used in an appropriate context and we note that the general issue of screening versus surveillance was already explicitly addressed in Section 4.3.

4.The EtD framework is incompletely reported, and a required dimension is missing. The Results section lists only brief judgment statements (e.g., "Judgment: Small," "Judgment: Low") without presenting the full GRADE EtD tables. Readers cannot trace how the working group weighed benefits, harms, certainty, patient values, cost-effectiveness, equity, acceptability, and feasibility to reach each recommendation. More critically, the "patient values and preferences" dimension—an obligatory element of the GRADE-EtD framework—is entirely absent. Given the controversies around asbestos surveillance (potential psychological harm, radiation risk, limited treatment options), patient perspectives are essential.

It is suggested to provide complete GRADE EtD tables (as supplementary material or in-text tables) documenting the evidence summary and judgment rationale for each dimension. Add the missing "values and preferences" domain, drawing on any available patient-preference literature or, if absent, a transparent expert consensus judgment with explicit reasoning.

Response 4

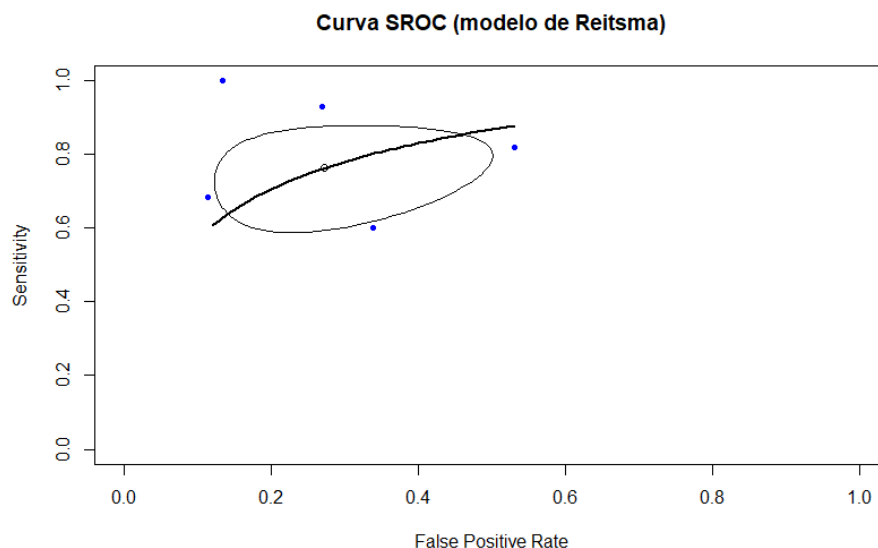
EtD development is an iterative process, and the latest version includes planetary health; however, this EtD framework was developed under the publication by Alonso-Coello P et al 2016, and the included criteria were according to Table 1 of this publication. The EtD framework will be included in the supplementary material (Appendix 7. Summary of EtD judgements). In terms of patient (and public) representation, FUNDCLAS (Association for Colombia Free of Asbestos) was part of the working group; this group represents, in part, patient values and preferences. In addition, Section 4.2 of the main manuscript ('Implementation considerations') mentions patient associations..

5. Diagnostic accuracy data are not quantitatively synthesized. Only five of the 23 included studies for Question 1 reported sensitivity and specificity. The authors report only wide ranges (sensitivity 0.60–0.94; specificity 0.47–0.88) without pooled estimates, confidence intervals, or formal assessment of heterogeneity sources. For diagnostic test accuracy studies, meta-analysis using bivariate or HSROC models is standard practice and is necessary to judge precision and consistency.

Recommendation: Perform a quantitative synthesis of the five diagnostic accuracy studies (e.g., using the *mada* package in R or the *midas* module in Stata), reporting pooled sensitivity/specificity with 95% CIs and an SROC curve. If pooling is deemed inappropriate due to excessive heterogeneity (e.g., differing reference standards, populations, or index test protocols), state this explicitly and present subgroup or sensitivity analyses to explore sources of variability.

Response 5

Thank you for your advice. We executed the analysis in R using the *mada* package and obtained the following plot, and we will add to the supplementary material; however, the plot showed considerable dispersion of studies, suggesting heterogeneity between studies. The summary point (~0.76 sensitivity, ~0.73 specificity) gives a reasonable overall estimate, but the ellipse is relatively large and horizontally elongated, indicating that the summary specificity has more uncertainty than the sensitivity. With only 5 studies, the bivariate confidence intervals are generally wide—this is a limitation of statistical power. In Line 245 the sentence “and the SROC curve are” was added.



6. The indirectness (applicability) of the included evidence to the Colombian context is insufficiently discussed. All included studies originate from high-income countries (e.g., Sweden, Canada, Australia, USA, Finland, Japan). Not a single study comes from Colombia or a comparable middle-income setting. Differences in asbestos fiber types, cumulative exposure levels, occupational versus environmental exposure patterns, prevalence of comorbidities (e.g., obesity, smoking), radiological infrastructure, and reader expertise may substantially affect the diagnostic performance of chest X-ray and the feasibility of the proposed surveillance model.

Response 6

Thank you for noting this. We commented on the indirectness of the evidence and decided to adjust Appendix 6 downgrade one level in the indirectness domain; the overall certainty changed from low to very low, and this change is in concordance with the conditional recommendation.

For Question 1, we included only 2 articles from Brazil out of a total of 23 articles, Brazil being the closest to the Colombian context. We added the sentence: “Another limitation was the indirectness of the evidence for Question 1, with only 2 out of 23 studies coming from Brazil, the closest context to Colombia; the remaining articles as well as those analyzed in Questions 2 and 3 came from high-income countries.”