Use of Artificial Intelligence for Rapid Epidemiology Reviews with Pooled Prevalence and Incidence Estimates

Allie Cichewicz[1], Kassandra Schaible[1]

[1]Thermo Fisher Scientific, Waltham, MA, USA

EPH41

Background

- Rare diseases impact more than 300 million people globally and more than 30 million in the United States (US) alone;[1,2] many of which have no approved treatments.
- Orphan drug designations for rare disease are often offered by regulatory agencies, if it can be demonstrated that the prevalence of disease falls within a specific threshold, defined by the US Food and Drug Administration (FDA) as fewer than 200,000 prevalent cases in the US.[1]
- Applications for orphan drug designations require the most relevant and recent values of disease prevalence. However, additional epidemiological data (e.g., incidence, mortality) are also useful to demonstrate unmet need in rare diseases.
- Comprehensive review of the existing published literature is helpful to identify and collate the required epidemiological data to support orphan drug applications; however, traditional systematic literature review (SLR) processes tend to be time consuming and may delay the application process if not prepared in advance.
- The use of artificial intelligence (AI) has been shown to expedite steps of the literature review process, but the capacity in which AI-driven rapid reviews can support these capabilities without sacrificing quality or robustness has yet to be established.

Objectives

- **We aimed to investigate whether a rapid review approach with AI integration was able to replicate the findings of recently published epidemiology SLRs with meta-analysis (MA) in two rare disease indications that have been granted orphan drug designations by the FDA: Takayasu arteritis (TAK) and non-cystic fibrosis bronchiectasis (NCFBE).**

Methods

- Published SLRs/MAs evaluating the epidemiology of TAK[3] and NCFBE[4] were replicated using a rapid reviews approach in Nested Knowledge (NK), leveraging AI capabilities to conduct the searches and screening.
- Based on each research question, Smart Search developed an individual algorithm to search PubMed for literature relevant to the incidence of TAK and prevalence of NCFBE, respectively. Of the three search options provided by Smart Search, the algorithm with the smallest search yield was selected.
- Two approaches were used for rapid screening:
  - For the TAK review, CORE smart tags were applied to the search yield and used to inform initial population, intervention, comparator, and outcome (PICO)-based screening to exclude studies deemed to be irrelevant (i.e., by study design) and to identify potential includes until a minimum of 50 records were screened to train the machine-learning model (robot screener).
  - For the NCFBE review, 50 records were initially screened by one human reviewer to arrive at the minimum training set required for robot screener.
- In both reviews, advancement probabilities were then leveraged for inclusion/exclusion decisions. Records with an advancement probability <0.2 were bulk excluded and an additional 50 records were screened. Only records with an advancement probability of ≥0.8 were included.

Methods (cont.)

- Studies included in these rapid reviews were cross-checked against those originally included in their respective published SLR. For all included records, data were manually extracted from the full-text documents within NK's meta-analytical extraction module and used as inputs to automatically generate pooled incidence and prevalence estimates.

Results

- The searches resulted in 971 and 914 records to be screened for TAK and NCFBE, respectively.
- AI-generated searches picked up approximately 90% of the included articles in the published SLRs/MAs. Studies that were not picked up by the Smart Search algorithms (n=3) were not indexed in PubMed.
- Leveraging AI tools for study selection resulted in rapid review inclusion of 54% (**Table 1**) to 73% (**Table 2**) of the articles included in the published SLRs/MAs.
  - For studies that were included in the published SLRs but excluded from the rapid reviews, most were excluded for study design, a result of bulk exclusion via smart tags.
- Therefore, two MA estimates were generated based on the rapid reviews and compared to the published MA estimates:
  1. **Recreated estimate**: only studies included in the published SLR/MA
  2. **Rapid estimate**: only studies that were included in the rapid review

Table 1. Included TAK Studies in the Published SLR (Rutter et al. 2021)[3]

Primary Author	Year	Country	Sample Size[a]	Incidence Rate Per 1,000,000 Persons	Eligibility Status
Kanecki[5]	2018	Poland	177	0.92	●
Gudbrandsson[6]	2017	Norway	78	1.50	✗
Makin[7]	2017	Australia	13	0.42	●
Park[8]	2017	South Korea	612	2.40	✗
Nesher[9]	2016	Israel	11	2.10	●
Saritas[10]	2016	Turkey	23	3.40	●
Birlik[11]	2015	Turkey	41	1.11	✗
Mohammed[12]	2015	Sweden	13	0.70	–
Romero-Gomez[13]	2015	Spain	5	1.10	–
Dreyer[14]	2011	Denmark	19	0.40	●
Watts[15]	2009	UK	14	0.80	●

Abbreviations: SLR = systematic literature review; TAK = Takayasu arteritis
Published SLR search executed in November 2019.
[a] Number of patients with TAK; ● included in rapid review; ✗ excluded from rapid review; – not identified by search

Table 2. Included NCFBE Studies in the Published SLR (Wang et al. 2024)[4]

Primary Author	Year	Country	Sample Size[a]	Incidence Rate Per 1,000,000 Persons	Eligibility Status
Gibbs[16]	2024	Australia	459	1,935	●[b]
Feng[17]	2022	China	383,926	101	●
Yang[18]	2022	South Korea	78	376	✗
Diaz[19]	2021	US	209	9,600	●
Kim[20]	2021	South Korea	1,005	1,546	●
Wu[21]	2020	China	NR	1,392	–
Yang[22]	2020	South Korea	376	393	●
Choi[23]	2019	South Korea	30,732	464	●
Diel[24]	2019	Germany	17,095	18	●
Weycker[25]	2017	US	31,122	94	●
Zhou[26]	2013	China	135	1,249	✗
Goeminne[27]	2012	Belgium	539	2,567	●
Seitz[28]	2012	US	22,296	1,103	●
Kwak[29]	2010	South Korea	1409	9,155	✗
Weycker[30]	2005	US	1,424	25	●

Abbreviations: NCFBE = non-cystic fibrosis bronchiectasis; SLR = systematic literature review
Published SLR search executed in May 2024.
[a] Number of patients with NCFBE; [b] Study was picked up by the AI but manually excluded at the full-text stage due to having a very specific population;
● included in rapid review; ✗ excluded from rapid review; – not identified by search

Results (cont.)

- The incidence of TAK was estimated to be 1.1 per 1,000,000 population in the previously published SLR/MA (published estimate). When estimating the incidence via rapid review (rapid estimate), the estimate was nearly identical (**Figure 1**). Prevalence estimates for NCFBE varied between the published estimate and the rapid estimate (680 vs. 570 per 100,000; **Figure 2**). However, some of the data included in the published SLR were for highly specific populations. Deletion of this data may lead to more generalizable results.
- Neither published estimate was perfectly regenerated when pooling only data from all originally included articles (recreated estimate).
  - Data from two of the included studies in the published NCFBE SLR could not be verified as a full-text document was not available for one of the publications and the other publication technically reported incidence and the previous authors of the SLR/MA included it as prevalence. Also, the weighting of estimates in the original article was unclear.
- The use of AI to aid in search development and screening, in addition to automation of MA, allowed for a rapid turnaround between search and analysis outputs, needing only around 15% of the time that would be required for a systematic approach.

Figure 1. Study-Level and MA Estimates for Incidence of TAK

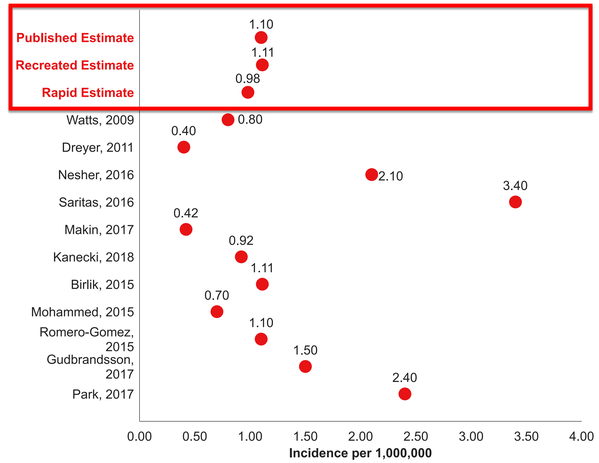
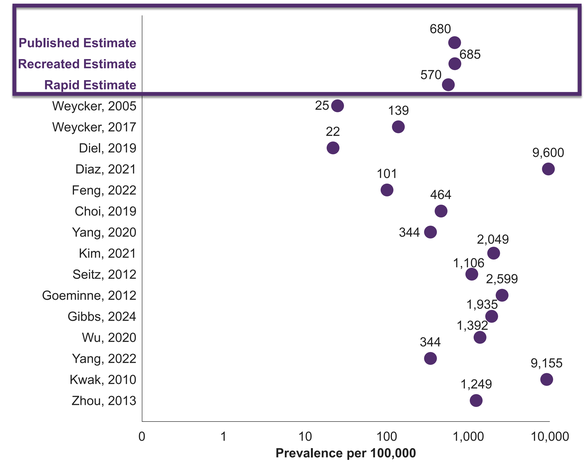


Figure 2. Study-Level and MA Estimates for Prevalence of NCFBE



Conclusions

- **AI and automation tools are crucial for quick evidence generation, and rapid reviews can arrive at similar study inclusion and findings to SLR/MAs where the evidence base was identified by more robust methodology.**
- **This approach may be beneficial for supporting orphan drug and other regulatory applications, which require evidence of low prevalence or high unmet need in the geographic areas of interest.**
- **Additional analyses of epidemiology estimates should be explored, as there can be heterogeneity in the way the data are collected and reported. Given this, it is important to consult with methodological experts and trained epidemiologists when submitting these data to regulatory authorities.**

References

1. Food and Drug Administration. Rare Diseases at FDA. 2024. https://www.fda.gov/patients/rare-diseases-fda; **2.** The Lancet Global Health. *Lancet Glob Health.* 2024;12(3):e341.; **3.** Rutter M, et al. *Rheumatology (Oxford).* 2021;60(11):4982-4990.; **4.** Wang L, et al. *BMC Public Health.* 2024;24(1):2675.; **5.** Kanecki K, et al. *Ann Agric Environ Med.* 2018;25(3):469-472.; **6.** Gudbrandsson B, et al. *Arthritis Care Res (Hoboken).* 2017;69(2):278-285.; **7.** Makin K, et al. *Mod Rheumatol.* 2017;27(6):1019-1023.; **8.** Park S, et al. *Int J Cardiol.* 2017;235:100-104.; **9.** Nesher G, et al. *J Rheumatol.* 2016;43(6):1072-7.; **10.** Saritas F, et al. *Rheumatol Int.* 2016;36(7):911-6.; **11.** Birlik M, et al. *Clin Exp Rheumatol.* 2016;34(3 Suppl 97):S33-9.; **12.** Mohammad A, et al. *J Rheumatol.* 2015;42(5):853-8.; **13.** Romero-Gomez C, et al. *Clin Exp Rheumatol.* 2015;33(2 Suppl 89):S-11-8.; **14.** Dreyer L, et al. *Clin Exp Rheumatol.* 2011;29(1 Suppl 64):S40-2.; **15.** Watts R, et al. *Rheumatology (Oxford).* 2009;48(8):1008-11.; **16.** Gibbs C, et al. *Med J Austin.* 2024;220(4):2020-95.; **17.** Feng J, et al. *Respiratory research.* 2022;23(1):111.; **18.** Yang B, et al. *J Pers Med.* 2022;12(5).; **19.** Diaz A, et al. *Radiology.* 2021;300(1):190-96.; **20.** Kim S, et al. *ERJ Open Res.* 2021;7(3).; **21.** Wu D, et al. *J Practical Med.* 2020;36(14):2003-06.; **22.** Yang B, et al. *Ann Transl Med.* 2020;8(21):1350.; **23.** Choi H, et al. *Eur Respir J.* 2019;54(2).; **24.** Diel R, et al. *Respir Med.* 2019;151:121-127.; **25.** Weycker D, et al. *Chronic resp disease.* 2017;14(4):377-384.; **26.** Zhou Y, et al. *Zhonghua Nei Ke Za Zhi.* 2013;52(5):379-82.; **27.** Goeminne P, et al. *Respir Res.* 2012;13(1):21.; **28.** Seitz A, et al. *Bench to Bedside Advances in Bronchiectasis.* 2011;A4869-A4869.; **29.** Kwak H, et al. *Tohoku J Exp Med.* 2010;222(4):237-42.; **30.** Weycker D, et al. *Clinical Pulmonary Medicine.* 2005;12(4):205 EP-209.

Disclosures

AC and KS are employees of PPD™ Evidera™ Health Economics & Market Access, Thermo Fisher Scientific. Funding provided by Thermo Fisher Scientific.

Acknowledgments

Editorial and graphic design support were provided by Michael Grossi and Karissa Calara of Thermo Fisher Scientific.