Evaluation of the fragility of pivotal trials used to support US Food and Drug Administration approval for plaque psoriasis

Sophia Z. Shalhout, PhD,[a] Romi Bloom, MD,[b] Lynn Drake, MD,[b,c] and David M. Miller, MD, PhD[a,b]
Boston, Massachusetts

Background: Over the last 5 years, there has been a rapid growth in the number of clinical trials used to support a US Food and Drug Administration (FDA) approval for systemic therapies with labeled indications for plaque psoriasis.

Objective: We aim to evaluate the fragility of clinical trial data used to support FDA approval of therapies for psoriasis.

Methods: We reviewed the primary endpoints of the pivotal trials of all systemic medications with a labeled indication for plaque psoriasis available from Drugs@FDA.

Results: Sixty-nine clinical trial primary endpoints met inclusion criteria and were assessed for robustness, yielding a median fragility index of 72 and a median fragility quotient of 0.19.

Limitations: Efficacy and statistical analysis data for several approved medications were not available on the product label or on Drugs@FDA.

Conclusions: When compared with randomized controlled trials for FDA approval across various diseases, pivotal trials in psoriasis appear quite robust to changes in outcomes. (J Am Acad Dermatol 2021;84:354-60.)

Key words: biologics; fragility index; fragility quotient; Physician Global Assessment; pivotal trials; psoriasis; Psoriasis Area and Severity Index.

Psoriasis is a widespread dermatologic condition affecting >8 million Americans. In recent years, there has been rapid growth in the number of agents approved by the US Food and Drug Administration (FDA) for psoriasis (Fig 1). The first therapy approved for psoriasis was methotrexate in 1972. There are currently a total of 28 distinct systemic therapies that have been approved by the FDA for the treatment of psoriasis. Strikingly, 64% of these therapies were approved in the last 5 years (Fig 1). Fueling the proliferation of approvals has been the discovery that specific effector molecules, such as tumor necrosis factor, interleukin-17, and interleukin-23 are critical in the pathogenesis. Commensurate with the diverse molecular origins, there are 12 distinct therapeutic classes of medications approved for psoriasis (Fig 1).

To demonstrate clinical benefit and support FDA approval to market a new drug, the proposed indication is typically evaluated by prespecified, primary endpoint(s) in pivotal trials assessing efficacy. Primary efficacy endpoints are the basis for the design and success of a trial, which is determined by significance testing of proposed hypotheses. To

From the Division of Hematology/Oncology,[a] Department of Dermatology,[b] and the Wellman Center for Photomedicine,[c] Massachusetts General Hospital, Harvard Medical School.
Funding sources: None.
Dr Miller has received honoraria for participating on advisory boards for Pfizer, Merck, Regeneron, and Sanofi Genzyme.
Accepted for publication April 7, 2020.
Reprints not available from the authors.

Address correspondence to David M. Miller, MD, PhD, FAAD, Division of Hematology/Oncology, Massachusetts General Hospital, Bartlett Hall, Rm 132, 15 Parkman St, Boston MA 02114. E-mail: dmiller4@mgh.harvard.edu.
Published online April 19, 2020.
0190-9622/$36.00

https://doi.org/10.1016/j.jaad.2020.04.057

assess the robustness of the randomized controlled trials (RCTs) used in therapies for psoriasis that are approved by the FDA, we used the fragility index (FI), an established metric for evaluating the statistical fragility of clinical trial data. The FI demonstrates the ease for which statistically significant results are lost with alterations in the numbers of the outcomes.[1] To calculate the FI of RCTs, the results of a trial are arranged in a 2 × 2 contingency table (Supplemental Fig 1 available via Mendeley at https://doi.org/10.17632/5sxhn6hn6g.1). The total number of subjects in the trial arm is maintained constant throughout each iteration of single step event status modifications. An event is added to the group with the smaller number of events, while subtracting a nonevent. If the new *P* value produced by a Fisher exact test does not equal or exceed .05, then another round of these modifications continues. If the new *P* value produced equals or exceeds .05, the number of events added to reach this *P* value is the FI. If the *P* value is still not .05 or greater, iterations continue until the first instance where the Fisher exact test equals .05 or greater. For example, a RCT with a FI of 3 indicates that only 3 subjects in that study would need to change from an event to a nonevent to alter the trial result from significant to nonsignificant. We provide an in-depth analysis of the robustness of pivotal trials used to support the FDA approvals of systemic therapies in psoriasis.

CAPSULE SUMMARY

- Fragility index is a metric used to assess the robustness of clinical trial data. The fragility index is the minimum number of subjects whose outcome would have to change from an event to a nonevent to alter the trial result from significant to nonsignificant.
- Pivotal trials in psoriasis are quite robust to changes in outcomes.
- Additional consideration regarding trial design with more emphasis on conservation of resources in psoriasis is appropriate.

METHODS

To evaluate the evidence used to support a labeled indication in psoriasis, this analysis includes new drug application or biological license application medications indexed on the FDA website (https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm). The selection criteria used in these analyses can be found in Supplemental Fig 2 (available via Mendeley). To date, 28 therapies have been approved by the FDA for the treatment of plaque psoriasis. Four of these therapies were not included in our analysis because the clinical trial data for the new drug application/biological license application were not available for review. Ten biosimilars were excluded from this analysis due to the fundamental differences required for the approval of biosimilars. A total of 33 efficacy trials were used to support an FDA approval in psoriasis by the remaining 14 therapies included in our analysis. We analyzed the predefined primary endpoints that were used to support the therapies' first label approved by the FDA with a psoriasis indication that met inclusion criteria for FI assessment (ie, dichotomous outcome, 2-parallel design). Sixty-nine primary or coprimary endpoints met these inclusion criteria. FIs were calculated using the package 'fragilityindex' and figures were generated with the package 'ggplot2,' using the R programming language, version 3.6.1 (R Foundation for Statistical Computing). The confidence level for the FI calculation was based on the α level used in pivotal trials indicated on the product label. For coprimary endpoints, the confidence level was adjusted according to the α splitting performed by the FDA. Fragility quotients (FQs) were calculated by dividing the FI by the number of subjects in that trial. Raw data for FI and FQ calculations can be viewed online (https://www.themillerlab.io/publication/fragility-of-pivotal-trials-in-psoriasis/).

RESULTS

The primary efficacy endpoints of pivotal trials used to support an FDA approval in psoriasis have been evaluated exclusively with the use of 2 instruments: Psoriasis Area and Severity Index (PASI) and Physician Global Assessment (PGA). Each trial was designed by comparing the efficacy of the proposed therapeutic with a placebo, with only 1 application incorporating an additional active comparator arm as a primary endpoint that was included on the label: the biological license application for brodalumab incorporated ustekinumab as a coprimary. Of note, other sponsors' development programs included active biologic comparator arms but were prespecified as secondary or exploratory endpoints and were not included in this analysis. All submitted applications had ≥2 pivotal trials and several used 3 pivotal trials to support their efficacy labeling claims in psoriasis for a total of 33 trials (Fig 2). The mean number of subjects enrolled in the arms prespecified for primary and coprimary endpoint

Abbreviations used:

FDA: U.S. Food and Drug Administration
FI: fragility index
FQ: fragility quotient
MFI: median fragility index
MFQ: median fragility quotient
PASI: Psoriasis Area and Severity Index
PGA: Physician Global Assessment
RCT: randomized controlled trial

analysis in the labeled pivotal trials was 1421 ± 575 (mean ± standard deviation). However, this average subject number underestimates the mean of the total number of subjects enrolled in pivotal trials because subjects are also enrolled in other arms prespecified for dose escalating studies or secondary endpoint analysis.

The FI was used to evaluate the robustness of the results reported in each primary endpoint hypothesis test. Because of fundamental differences in the approval process for biosimilars, only data from nonbiosimilars were included in the analysis of primary endpoints. The median FI (MFI) of pivotal trials in psoriasis was 72 (Fig 3). The FQ is a statistical summary used to adjust for the potential effects on FI from variations in study subject number size. To normalize for sample size differences across trials, the FQ was evaluated for each of the clinical trial primary endpoint results in psoriasis, yielding a median FQ (MFQ) of 0.19 (Fig 4).

Many of the 14 psoriasis systemic therapies also carry an FDA approval for the treatment of nonplaque psoriasis indications as well, including rheumatoid arthritis, Crohn's disease, and psoriatic arthritis. In comparison to the MFI of 72 in psoriasis, the MFI of pivotal trial primary endpoints used for the FDA approval of nonplaque psoriasis indications was only 15, approximately 5-fold less than the MFI for trials in psoriasis (Fig 5). The MFQ for the nonplaque psoriasis trials was 0.077, approximately 2.5-fold less than the MFQ for trials in psoriasis (Supplemental Table I available via Mendeley).

DISCUSSION

There has been a rapid proliferation of medications approved by the FDA with labeled indications for psoriasis over the last decade, with particular growth over the last 5 years. The reasons are likely multifactorial, involving advances in the development of biosimilars and perhaps, most importantly, improvement in our understanding of the pathophysiology of psoriasis that has led to the development of effective targeted treatments. We evaluated the statistical persuasiveness of the efficacy pivotal trial endpoints used in the FDA drug approval process to establish clinical benefit in therapies with an indication for plaque psoriasis.

In their seminal paper, Walsh et al[1] evaluated 399 RCTs across several diseases and found the MFI to be 8. Since that publication, the FI of trials in numerous medical conditions has been assessed. We performed a literature search and review of previous articles where the FIs of various diseases and clinical trials in medicine were evaluated.[2-5] No articles assessing the FI in any psoriasis trials were found in the literature search. Our analysis included 49 previous reports encompassing 3632 trials revealing an overall low MFI of 3 (Supplemental Tables II and III available via Mendeley). This is in sharp contrast to an MFI of 72 in the psoriasis pivotal trials reported herein. A caveat in this comparison is that we restricted our analysis to psoriasis efficacy pivotal trials used to support FDA-approved labeled claims, whereas many trials in our analysis of the FI in the literature focus on earlier exploratory RCTs. Given that later-phase trials benefit from information gained from earlier-phase investigations, earlier studies may be more fragile. That said, a recent evaluation of the robustness of pivotal trials used to support FDA approvals in oncology found the MFI to be 2.[5] However, oncologic indications appropriately have different standards for approval.

Interestingly, when we restricted a separate analysis of the MFI and MFQ to the labeled pivotal trials used for FDA approval in other nondermatologic inflammatory conditions of the same psoriasis medications, using the exact same strict inclusion criteria, the MFI was 5-fold lower than in psoriasis trials and the MFQ was 2.5-fold lower. Although comparisons with other fields are limited by salient differences in patient population, disease incidence, and severity, there may be some utility in the juxtaposition. That notwithstanding, an evaluation of primary endpoints of efficacy pivotal trials for FDA approvals in atopic dermatitis (dupilumab, MFI of 41), hidradenitis suppurativa (adalimumab, MFI of 20.5), and pemphigus vulgaris (rituximab, MFI of 16) revealed substantively lower MFIs compared with psoriasis (Supplemental Table IV available via Mendeley). Admittedly, cross-disciplinary comparisons must be made cautiously, but these comparisons reveal the robustness of the clinical trials in psoriasis are not simply explained because of the advantages of later phase trials, the criteria needed for FDA approval, or the selection criteria for analysis. Thus, when compared with RCTs for FDA approvals across various diseases, pivotal trials in psoriasis appear quite robust to changes in outcomes.

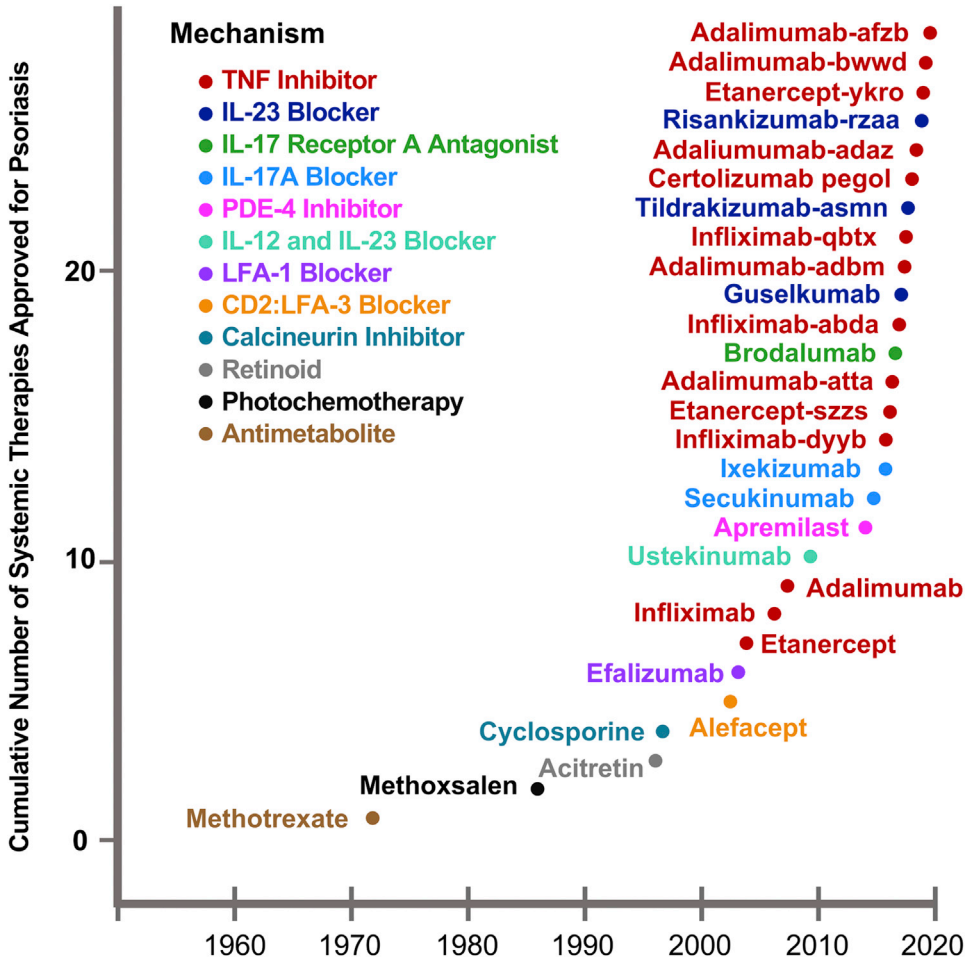

Fig 1. Psoriasis therapy time series showing the exponential growth in the development of biologics within the past 5 years. *CD2*, Cluster of differentiation 2; *IL*, interleukin; *LFA*, lymphocyte function-associated antigen; *PDE-4*, phosphodiesterase type 4; *TNF*, tumor necrosis factor.

There is no current agreement on what is the most appropriate FI; consequently, interpretation of its meaning is prone to subjectivity. In addition, there are valid criticisms regarding the use of FI and FQ in evaluating RCTs, especially when low FIs are used to criticize the strength of trial conclusions.[6,7] Acuna et al[7] argue that in an effort to minimize risk to patients and optimize resource use, RCTs are purposefully planned to enroll the least number of subjects required to detect a minimal meaningful clinical benefit. Therefore, it is not surprising, and maybe even appropriate, that many RCTs demonstrate a degree of fragility. However, the RCTs used in efficacy pivotal trials for indications in psoriasis are exceptionally robust.

While there is no specific standard level of robustness for pivotal trials, a post hoc power analysis of the pivotal trials in psoriasis revealed a median power of 100% for the primary endpoints (Supplemental Table I available via Mendeley). Indeed, assuming the exact same event rate in the psoriasis pivotal trials but reducing the number of subjects by 50% still results in a robust MFI of 33 for all the development programs, with none losing the ability to reject the null. One must consider that in addition to efficacy, long-term safety is a critical consideration in the number of subjects enrolled in trials. Nevertheless, in the era of rising drug costs and therapeutics with generally acceptable risk profiles and statistically persuasive efficacy, perhaps additional consideration regarding trial design with more emphasis on conservation of resources in psoriasis is appropriate.

For example, to conserve resources in pivotal clinical trials in psoriasis, the use of coprimary endpoints in psoriasis may be revisited. There are numerous clinical measures used to evaluate psoriasis in routine practice and in clinical trials, though

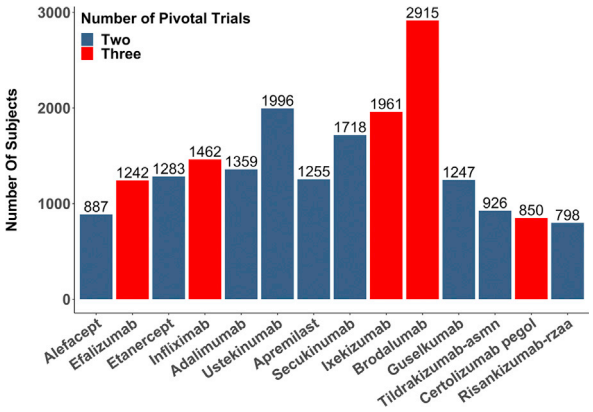

Fig 2. Total number of subjects in the primary and coprimary endpoint trial arms used in pivotal trials that were evaluated by the US Food and Drug Administration for approval in psoriasis. These include only the trial arms prespecified for primary endpoint analysis in the labeled pivotal trials; subjects enrolled in other arms of the pivotal trials prespecified for dose escalating studies or secondary endpoint analysis are excluded.

no one superior tool has emerged.[8] Puzenat et al[9] performed a systemic review of clinical studies in psoriasis and concluded that the PASI instrument was the most complete and most extensively validated score. As a result of its performance, improvements in PASI have been used as a primary endpoint by every nonbiosimilar application in our analysis since its inception with the exception of one. Nevertheless, because of limitations in assessment of certain aspects of the disease (eg, cases of minimal body surface area involvement or involvement of acral skin), PGA has been routinely incorporated as a coprimary endpoint in drug development programs, including every approved application since 2015. Although important in possibly minimizing the limitations of any one assessment tool, the use of coprimary endpoints affects clinical trial design because α splitting to conserve a familywise error rate can necessitate an inflation in sample size. However, the notion that PASI and PGA are sufficiently complementary to justify mandated use as coprimary endpoints has recently been challenged. After analyzing 30 RCTs in patients with moderate-to-severe psoriasis, Robinson et al[10] concluded that PASI and PGA had substantial redundancy for agents that produced a 75% reduction in PASI scores in ≥25% of patients. Indeed, in the 9 applications from our analysis that used both PASI 75 and PGA as coprimary endpoints in their pivotal trials, PASI 75 and PGA were significantly correlated (Pearson correlation 0.75, P = .00002; Supplemental Fig 3 available via Mendeley). The use of coprimary endpoints has implications on the cost of drug development programs and patient risk because it affects the number of subjects needed to establish statistical significance and therefore affects the number of patients exposed to unproven and potentially harmful investigational agents. Indeed, Robinson et al[10] argue that because of the divergence of PASI and PGA at lower therapeutic effectiveness, coprimary endpoints may be more appropriate for early-phase clinical studies. In contrast, with efficacy pivotal trials, where more clarity regarding efficacy has already been established in earlier parts of a drug development program, consideration for a single primary endpoint with PASI may be warranted to conserve costs and resources.

Furthermore, current legislative statutes do not have an active comparator efficacy requirement; thus, sponsors do not need to provide evidence that the investigational agent is superior to currently available therapy. Therefore, the recent incorporation of active comparators into trial design is not mandated by statute but increases cost and resources necessary for trials. However, this incorporation of active comparator arms may reflect commercial strategic factors, such as insurance formulary preference or fulfillment of a global development program. In addition, the inclusion of active comparator arms may be ethically favored when designed to compare the new therapy against the best available standard of care.

To provide "substantial evidence" of clinical benefit, the null hypothesis of a pivotal trial's primary endpoint must be rejected. To do so, there must be statistical persuasiveness that the effect seen is not caused by chance. The probability of rejecting the null hypothesis is related to the power of a study, which is influenced by the effect size of the intervention, the decided upon significance level, and the sample size. Our analysis shows that the efficacy pivotal trials for systemic therapies with labeled indications for psoriasis appear to be substantially robust compared with those for other indications with an FDA approval. In the past, the FI has generally been used to indicate that most reported trials with statistical significance would lose this significance with a small change in the number of outcomes.[1-6,11-13] We show the opposite for the pivotal trials in psoriasis that are far less fragile and perhaps warrant consideration in trial design to reduce costs.

We thank Hang Lee, PhD, of the Massachusetts General Hospital—Harvard Catalyst Biostatistics Program for his assistance and guidance with the fragility index and the statistical analysis plan. We also thank Melissa Reyes, MD, MPH, DTMH, LCDR, USPHS, of the Division of

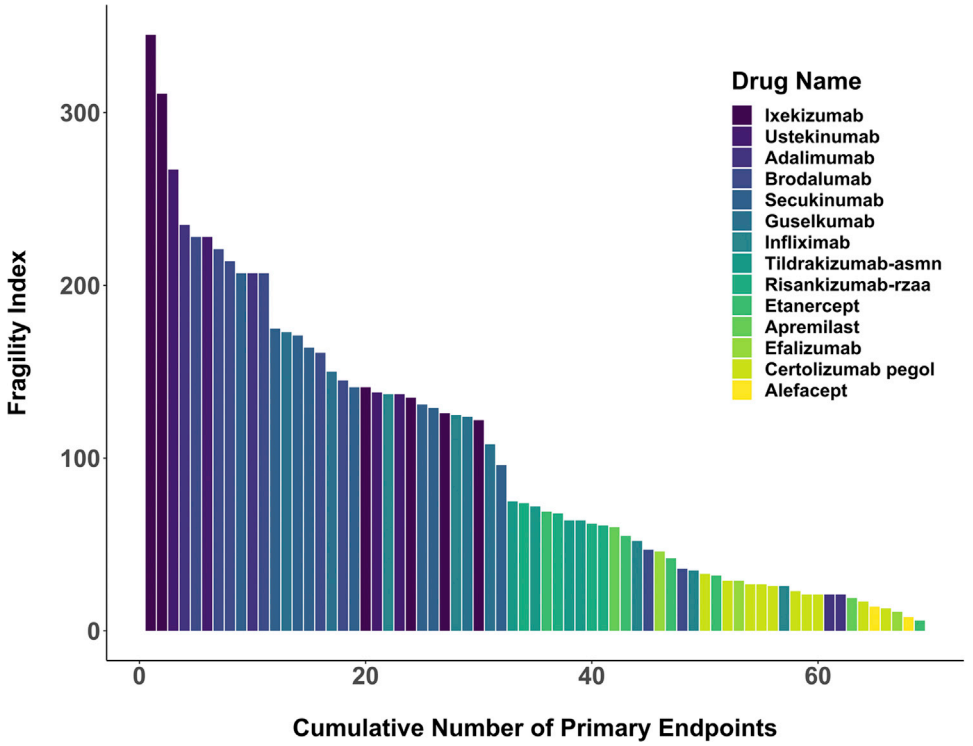

Fig 3. Fragility index of pivotal trial primary endpoints in plaque psoriasis. The fragility index was calculated for the primary or coprimary endpoints from 14 psoriasis therapies approved by the US Food and Drug Administration with available data. Sixty-nine endpoints met fragility index inclusion criteria for analysis. The median fragility index was 72.

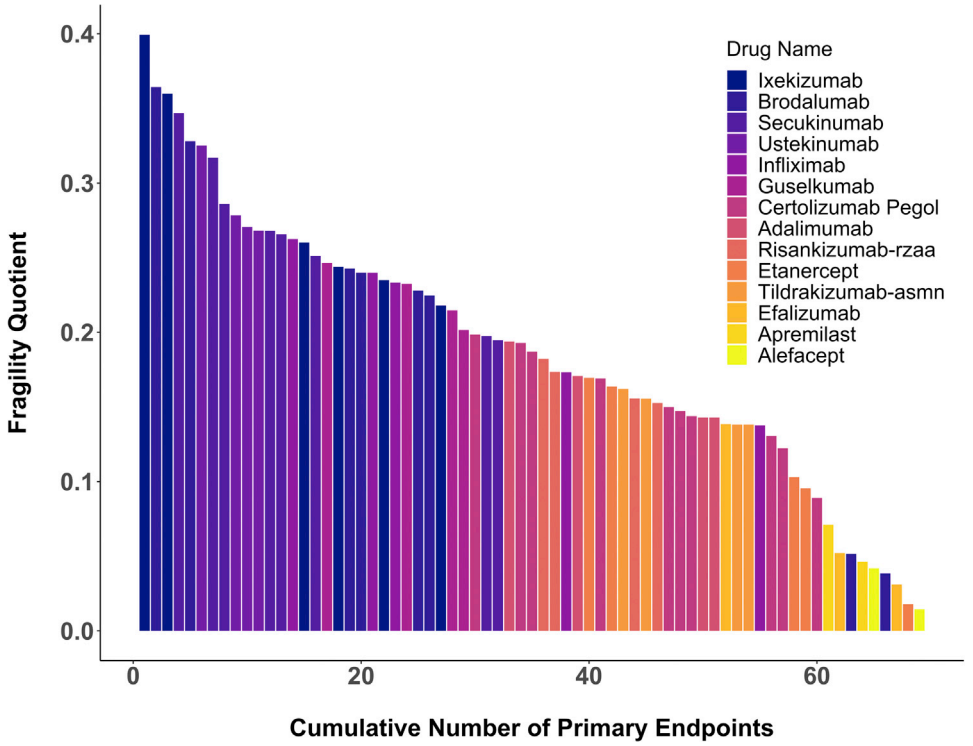

Fig 4. Fragility quotient of pivotal trial primary endpoints in plaque psoriasis. The fragility quotient was calculated for the 69 primary or coprimary endpoints that met fragility index inclusion criteria for analysis. The fragility quotient was determined by dividing the fragility index by the total subjects enrolled in the primary or coprimary endpoint arms. The median fragility quotient was 0.19.

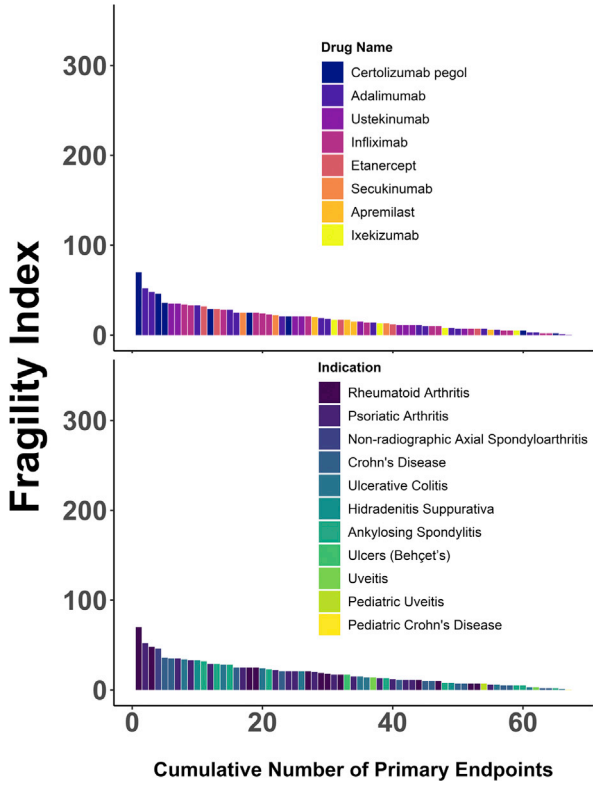

Fig 5. Fragility index of pivotal trial primary endpoints in nonplaque psoriasis indications. Eight of the biologic therapies with indications in psoriasis also carry approval by the US Food and Drug Administration for nonplaque psoriasis indications (eg, rheumatoid arthritis, Crohn's disease). The fragility index of the nonplaque psoriasis pivotal trial primary endpoints used for FDA approval and that met the selection criteria for fragility index analysis are graphed. The upper panel is color coded by drug name and the bottom panel is the corresponding fragility index data color coded by indication. The median fragility index is 15.

Dermatology and Dental Products, US Food and Drug Administration for her review and assistance. This article reflects the views of the authors and should not be construed to represent the views or policies of the US Food and Drug Administration.

REFERENCES

1. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol*. 2014;67:622-628.
2. Van Howe RS. The fragility index in HIV/AIDS trials [e-pub ahead of print]. *J Gen Intern Med*. https://doi.org/10.1007/s11606-019-05554-x. Accessed April 23, 2020.
3. Wayant C, Meyer C, Gupton R, Som M, Baker D, Vassar M. The fragility index in a cohort of HIV/AIDS randomized controlled trials. *J Gen Intern Med*. 2019;34:1236-1243.
4. Berti A, Cornec D, Medina Inojosa JR, Matteson EL, Murad MH. Treatments for giant cell arteritis: meta-analysis and assessment of estimates reliability using the fragility index. *Semin Arthritis Rheum*. 2018;48:77-82.
5. Del Paggio JC, Tannock IF. The fragility of phase 3 trials supporting FDA-approved anticancer medicines: a retrospective analysis. *Lancet Oncol*. 2019;20:1065-1069.
6. Niforatos JD, Zheutlin AR, Chaitoff A, Pescatore RM. The fragility index of practice changing clinical trials is low and highly correlated with P-values. *J Clin Epidemiol*. 2020;119:140-142.
7. Acuna SA, Sue-Chue-Lam C, Dossa F. The fragility index-P values reimagined, flaws and all. *JAMA Surg*. 2019;154:674.
8. Spuls PI, Lecluse LL, Poulsen MLNF, Bos JD, Stern RS, Nijsten T. How good are clinical severity and outcome measures for psoriasis?: quantitative evaluation in a systematic review. *J Invest Dermatol*. 2010;130:933-943.
9. Puzenat E, Bronsard V, Prey S, et al. What are the best outcome measures for assessing plaque psoriasis severity? A systematic review of the literature. *J Eur Acad Dermatol Venereol*. 2010;24(suppl 2):10-16.
10. Robinson A, Kardos M, Kimball AB. Physician Global Assessment (PGA) and Psoriasis Area and Severity Index (PASI): why do both? A systematic analysis of randomized controlled trials of biologic agents for moderate to severe plaque psoriasis. *J Am Acad Dermatol*. 2012;66:369-375.
11. Atal I, Porcher R, Boutron I, Ravaud P. The statistical significance of meta-analyses is frequently fragile: definition of a fragility index for meta-analyses. *J Clin Epidemiol*. 2019;111:32-40.
12. Meyer C, Bowers A, Tritz D, et al. The fragility of randomized trial outcomes underlying management of dyspepsia and *Helicobacter pylori* infections. *Int J Evid Based Healthc*. 2020;18:125-137.
13. Vargas M, Buonanno P, Iacovazzo C, Servillo G. Epinephrine for out of hospital cardiac arrest: a systematic review and meta-analysis of randomized controlled trials. *Resuscitation*. 2019;136:54-60.