Arthritis & Rheumatology
Vol. 71, No. 8, August 2019, pp 1225–1231
DOI 10.1002/art.40870

American College of Rheumatology
Empowering Rheumatology Professionals

Derivation and Validation of a Major Toxicity Risk Score Among Nonsteroidal Antiinflammatory Drug Users Based on Data From a Randomized Controlled Trial

Daniel H. Solomon,[1] Ming Shao,[2] Kathy Wolski,[2] Steven Nissen,[2] M. Elaine Husni,[2] and Nina Paynter[1]

Objective. While nonsteroidal antiinflammatory drugs (NSAIDs) are commonly used in rheumatology, they can cause major toxicity. Improving the risk/benefit ratio requires a more precise understanding of risk. This study was undertaken to derive and validate a risk score for major toxicity among NSAID users enrolled in a randomized controlled trial.

Methods. Patients enrolled in a randomized controlled trial who had known cardiovascular disease or risk factors as well as osteoarthritis or rheumatoid arthritis were divided into derivation and validation cohorts. Patients were randomized to receive celecoxib, naproxen, or ibuprofen at typical dosages. The risk score was designed to predict the 1-year occurrence of major toxicity among NSAID users, including major adverse cardiovascular events, acute kidney injury, significant gastrointestinal events, and mortality. Variables significantly associated with major toxicity were candidates for inclusion in the final regression model. After derived models were found to have a similar model fit in the validation set, the cohorts were combined, allowing calculation of a risk score.

Results. In the derivation cohort, significant variables included age, male sex, history of cardiovascular disease, hypertension, diabetes mellitus, tobacco use, statin use, elevated serum creatinine level, hematocrit level, and type of arthritis. The C-index was 0.73 in the validation cohort and 0.71 in the total cohort; the model was well calibrated. Of the total population with complete data (n = 23,735), 1,080 participants (4.6%) had a predicted 1-year risk of major toxicity of <1%, 16,273 (68.6%) had a predicted risk of 1–4%, and 6,382 (26.9%) had a predicted risk of >4%.

Conclusion. The risk score accurately categorizes the 1-year risk of major toxicity among NSAID users and may be useful in identifying patients who can safely use these agents.

INTRODUCTION

Nonsteroidal antiinflammatory drugs (NSAIDs) are the most common medications used regularly in the US, with estimates of 10 million people taking them daily (1). They are relatively safe treatment options when used intermittently by most people, prompting their widespread availability as over-the-counter preparations. However, among patients with comorbid conditions who use them at moderate dosages for long periods of time, they pose a risk of significant toxicity. Estimates from the late 1990s suggest that ~17,000 people died annually from NSAID toxicities (2). Mortality from NSAIDs may have decreased since the introduction of proton-pump inhibitors (3). Well-recognized toxicities include gastrointestinal (GI) complications, acute kidney injury, cardiovascular (CV) events, and death (4).

Balancing the analgesic and antiinflammatory benefits of NSAIDs with the potential for risk requires an appreciation of how clinical factors impact the relative and absolute risk of toxicity in a given patient. While genomics may lead us to a precision medicine–based approach to the use of these agents, basing recommendations on clinical risk factor data that is easily obtained at the individual level may help to improve their safety. Risk scores can facilitate application of risk factor epidemiology in the clinic (5). Good examples of this include the Framingham Risk Score (6), the Systematic Coronary Risk Evaluation (7), the American College of Cardiology/American

ClinicalTrials.gov identifier: NCT00346216.
The PRECISION trial was supported by Pfizer.
[1]Daniel H. Solomon, MD, MPH, Nina Paynter, PhD: Brigham and Women's Hospital, Boston, Massachusetts; [2]Ming Shao, PhD, Kathy Wolski, MPH, Steven Nissen, MD, M. Elaine Husni, MD, MPH: Cleveland Clinic Foundation, Cleveland, Ohio.
Dr. Shao and Ms Wolski have received research support from Pfizer. Dr. Nissen has received consulting fees and/or honoraria from Pfizer (less than $10,000) and research support from Pfizer paid to his institution. No other disclosures relevant to this article were reported.
Address correspondence to Daniel H. Solomon, MD, MPH, Brigham and Women's Hospital, Division of Rheumatology, 60 Fenwood Road, Boston, MA 02115. E-mail: dsolomon@bwh.harvard.edu.
Submitted for publication January 11, 2019; accepted in revised form February 21, 2019.

Heart Association guidelines (8), and the fracture risk assessment tool (9).

Risk score development and validation requires a large and accurately phenotyped cohort with thorough follow-up. We have applied recommended methods to derive and internally validate a risk score for major NSAID toxicity using data from the Prospective Randomized Evaluation of Celecoxib Integrated Safety versus Ibuprofen or Naproxen (PRECISION) trial (4,5).

PATIENTS AND METHODS

Study design and participants. The current study used information derived from the PRECISION trial, a large randomized controlled trial that compared the safety of celecoxib, naproxen, and ibuprofen (4). No placebo group was included. The trial was performed in 14 countries between 2006 and 2016. Local human subject committees approved the protocol, and all participants provided written informed consent that included use of data in secondary analyses. These secondary analyses were not prespecified in the statistical analysis plan. The trial sponsor had no authorship in the current manuscript but was given the opportunity to provide feedback.

The study population included patients with a clinical diagnosis of osteoarthritis (OA) or rheumatoid arthritis (RA) who were ≥18 years of age and required regular daily treatment with an NSAID. All participants were required to have a history of a CV event, occlusive coronary or noncoronary arterial disease, diabetes mellitus, or ≥2 CV risk factors for women and ≥3 for men. Risk factors included age ≥65 years for women and >55 years for men, hypertension, dyslipidemia, microalbuminuria, urine protein:creatinine ratio >2, ankle brachial index <0.9, left ventricular hypertrophy, current cigarette smoking, waist:hip ratio ≥0.90, and family history of premature CV disease (see Supplementary Table 1, on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40870/abstract). Exclusion criteria included a CV event within the prior 3 months, New York Heart Association class III or IV heart failure, and other characteristics listed in Supplementary Table 1. The instructions for completing the case report forms prompted study staff to use medication information and/or ask the patient each question.

All patients in the trial cohort were considered for inclusion in the analyses. Fewer than 1% of patients (n = 215) had some missing data at baseline and were excluded. The full cohort was divided into derivation and validation cohorts. As recommended (5), the split was not random; rather, patients enrolled in the first 4 years of the trial were included in the derivation cohort and those enrolled in the last 5 years were included in the validation cohort.

Outcome measures. The primary outcome measure for the current analyses was a composite of major toxicity among NSAID users. All toxicities were centrally adjudicated in a blinded manner and included major adverse CV events (MACEs), clinically significant GI events (CSGIEs), acute kidney injury, and death. MACEs included events defined by Antiplatelet Trialists Collaboration, in addition to revascularization or hospitalization for transient ischemic attack or unstable angina. CSGIEs included gastroduodenal hemorrhage, gastric outlet obstruction, perforation of the gastroduodenum or small/large bowel, hemorrhage of the small/large bowel, acute GI hemorrhage of unknown origin, or symptomatic gastric or duodenal ulcer. Acute kidney injury was defined as development of renal insufficiency or renal failure, including any of the following: serum creatinine level of ≥2.0 mg/dl and increase of ≥0.7 mg/dl from baseline, hospitalization for acute renal failure with either a doubling of the baseline serum creatinine level or hyperkalemia with ≥50% elevation in serum creatinine level, or initiation of dialysis. The composite primary outcome measure included all causes of death.

Examination for the occurrence of these outcomes took place during the first 1-year study period, while patients were receiving the study NSAID and for 30 days after termination of the NSAID, to capture only on-drug outcomes. Patients were censored from the analysis at the first instance of any of the following: death, after 1 year of receiving the study NSAID, or 30 days after discontinuing the study NSAID.

Potential risk factors. The potential predictors of major toxicity among NSAID users included variables measured at baseline that had been considered as potential risk factors in prior studies (2,10–12). We also focused on easily measured variables that could be considered in typical clinical practice. However, because we assumed that the variables would be weighted in the final risk score, requiring computer calculation, variables were not simplified into categories but left as continuous whenever possible.

Demographic variables included age and sex. Clinical variables included type of arthritis (OA or RA), known CV disease (i.e., prior event, occlusive coronary/noncoronary arterial disease), known diabetes mellitus, known hypertension, known hyperlipidemia, or a prior GI bleed. Medication use of interest included low-dose aspirin for CV prevention, statins/lipid-lowering drugs, or glucocorticoids (any dosage). We also considered a patient's functional status as measured with the Health Assessment Questionnaire (13), current tobacco use, and several laboratory measures (including serum creatinine and hematocrit levels).

Statistical analysis. After dividing the trial cohort into the derivation group (enrollment 2006–2010) and the validation group (enrollment 2011–2016), we examined baseline characteristics. Major toxicity rates among patients receiving NSAIDs were estimated across both groups. We then used data from the derivation cohort to fit a multivariable model using Cox proportional hazards regression, considering only the first toxicity event and a time period of days since randomization (PROC PHREG in SAS). Censoring events included a toxicity event, loss to follow-up, or 1 year, whichever occurred first. Age- and sex-

adjusted models were examined first. Each potential variable was tested in regression models. Variables with hazard ratios (HRs) of >1.2 or *P* values of <0.001 were considered potential risk factors and tested as a group in multivariable models, including age and sex. Interquartile ranges (IQRs) and 95% confidence intervals (95% CIs) were calculated. Hematocrit was tested using linear splines with a knot set at a hematocrit level of 43%. To construct a relatively parsimonious model, variables were removed if the HRs were no longer >1.2 in the multivariable model. Model discrimination was assessed using Harrell's C-index, and model calibration was examined by visually graphing observed and predicted risk curves with the calibration intercept and slope. The derivation cohort model was then applied to the validation cohort using the macros developed by Cook et al (14), and performance statistics were again assessed. All analyses were conducted using SAS (version 9.4) and the open-source software R.

Based on good performance in the validation cohort, the cohorts were combined for increased power and the risk model was reestimated using the same risk factors. The apparent performance of this final model in the full cohort was also assessed. Using the final risk model, a formula was developed (14) to predict risk probabilities in future patients. To address clinical decision making, 3 categories of 1-year risk of major toxicity were defined: low risk (<1%), intermediate risk (1–4%), and high risk (>4%). Survival curves were plotted for the 3 categories and risk of outcomes were compared using a log rank test (PROC LIFETEST in SAS). Finally, we assessed model calibration among several subgroups, by treatment arm and dosage. The 3 treatment arms included celecoxib, naproxen, and ibuprofen. The dosage analysis examined patients within each treatment group separately according to dosage, i.e., for each treatment, those who continued to receive the starting dosage (celecoxib 100 mg twice daily, naproxen 375 mg twice daily, or ibuprofen 600 mg 3 times daily) were assessed as a subgroup, and those in whom the dosage was up-titrated to improve analgesia (celecoxib 200 mg twice daily, naproxen 500 mg twice daily, or ibuprofen 800 mg 3 times daily) were assessed as another subgroup. Local drug labeling allowed the up-titration of celecoxib to 200 mg twice daily only for patients with RA.

RESULTS

The populations included in the cohorts to derive and validate the risk score were similar (Table 1). The median age in the combined cohort was 63 years (IQR 57–70), 64% were women, and the median body mass index was 31.4 kg/m^2 (IQR 27.5–36.3).

Table 1. Patient characteristics*

	Full cohort (n = 23,950)	Derivation cohort (n = 15,194)	Validation cohort (n = 8,756)
Age, median (IQR) years	63.0 (57.0–70.0)	62.0 (57.0–69.0)	65.0 (58.0–70.0)
Male sex	8,591 (35.9)	4,668 (30.7)	3,923 (44.8)
BMI, median (IQR) kg/m^2	31.4 (27.5–36.3)	31.3 (27.4–36.2)	31.6 (27.6–36.5)
Tobacco use	4,975 (20.8)	2,701 (17.8)	2,274 (26.0)
Type of arthritis			
OA	21,525 (89.9)	13,413 (88.3)	8,112 (92.6)
RA	2,425 (10.1)	1,781 (11.7)	644 (7.4)
History of diabetes mellitus	8,445 (35.5)	5,360 (35.7)	3,085 (35.2)
History of hypertension	18,644 (78.4)	11,617 (77.4)	7,027 (80.3)
History of hyperlipidemia	14,971 (63.0)	10,160 (67.7)	4,811 (55.0)
Prior CV event	1,201 (5.1)	653 (4.3)	548 (6.3)
Prior GI bleed	0 (0)	0 (0)	0 (0)
Serum creatinine, median (IQR) mg/dl	0.87 (0.74–1.02)	0.87 (0.74–1.02)	0.87 (0.73–1.01)
Hematocrit, median (IQR) %	41.0 (39.0–44.0)	41.0 (39.0–44.0)	42.0 (39.0–45.0)
Hematocrit level <43%	14,658 (61.2)	9,681 (63.7)	4,977 (56.8)
Use of aspirin	11,017 (46.0)	6,740 (44.4)	4,277 (48.8)
Use of statins	12,913 (53.9)	7,951 (52.3)	4,962 (56.7)
Use of glucocorticoids	3,089 (12.9)	1,914 (12.6)	1,175 (13.4)
Use of DMARDs	1,748 (7.3)	1,314 (8.6)	434 (5.0)
Functional status, median (IQR)†	1.1 (0.6–1.5)	1.1 (0.6–1.5)	1.1 (0.6–1.5)

* Except where indicated otherwise, values are number (%) of patients. IQR = interquartile range; BMI = body mass index; OA = osteoarthritis; RA = rheumatoid arthritis; CV = cardiovascular; GI = gastrointestinal; DMARDs = disease-modifying antirheumatic drugs.
† Measured with the Health Assessment Questionnaire, range 0 (no limitation) to 3 (unable to perform).

Table 2. One-year outcome rates for major toxicity among NSAID users*

	Full cohort (n = 23,950)			Derivation cohort (n = 15,194)			Validation cohort (n = 8,756)		
	Events	Person-years	Rate (95% CI)	Events	Person-years	Rate (95% CI)	Events	Person-years	Rate (95% CI)
Primary outcome measure	617	18,261	3.38 (3.12–3.65)	390	11,527	3.38 (3.06–3.72)	227	6,734	3.37 (2.94–3.81)
MACE	374	18,296	2.04 (1.83–2.25)	234	11,550	2.03 (1.76–2.29)	140	6,746	2.08 (1.74–2.41)
CSGIE	133	18,339	0.73 (0.60–0.85)	89	11,581	0.77 (0.60–0.93)	44	6,758	0.65 (0.45–0.84)
Acute kidney injury†	89	18,348	0.49 (0.38–0.59)	57	11,586	0.49 (0.37–0.62)	32	6,762	0.47 (0.31–0.63)
Death‡	90	18,354	0.49 (0.39–0.59)	52	11,591	0.45 (0.33–0.58)	38	6,763	0.56 (0.39–0.74)

* Rates are per 100 person-years. The primary outcome measure is a composite of major toxicities among nonsteroidal antiinflammatory drug (NSAID) users, which includes 4 components: major adverse cardiovascular event (MACE), clinically significant gastrointestinal event (CSGIE), acute kidney injury, and death. 95% CI = 95% confidence interval.
† Acute kidney injury was defined as development of renal insufficiency or renal failure, including any of the following: serum creatinine level of ≥2.0 mg/dl and increase of ≥0.7 mg/dl from baseline, hospitalization for acute renal failure with a doubling of the baseline serum creatinine level or hyperkalemia with ≥50% elevation in serum creatinine, or initiation of dialysis.
‡ Includes all causes.

Ninety percent of patients were diagnosed as having OA and 10% as having RA. Twenty-two percent had a known prior CV event, 36% had diabetes mellitus, 78% had hypertension, and 63% had hyperlipidemia. Patients were followed up for outcome measures 1 year after randomization, with 59% having ≥12 months of follow-up while receiving an NSAID. Of those with <12 months of follow-up while receiving an NSAID, the median follow-up was 3.7 months. Table 2 shows the incidence rates for the components and the composite of major toxicity among NSAID users in both cohorts. The rate of development of the primary outcome measure in the total cohort was 3.38% (95% CI 3.12–3.65).

To derive the risk score, we treated all variables from Table 1 as potential predictors of major toxicity among NSAID users. Variables were individually tested in age- and sex-adjusted models, and those with HRs of >1.2 or *P* values of <0.001 were included in a combined model. The final model was further pruned to include only variables with 95% CIs that excluded 1.00 (Table 3). The multivariable model fit statistics for the derivation cohort included a Harrell's C-index of 0.73 and near complete agreement in models examining observed risk in relation to predicted risk (Supplementary Figure 1A, http://onlinelibrary.wiley.com/doi/10.1002/art.40870/abstract). The model fit statistics in the validation cohort were similar to those in the derivation cohort, with a Harrell's C-index of 0.68 and a calibration slope of 0.774 (Supplementary Figure 1B).

After deriving the multivariable risk model and assessing its validity, we used the total cohort to calculate a risk score: (0.0325 × age) + (0.2666 × sex [male = 1]) + (0.8352 × known CV disease [yes = 1]) + (0.2252 × known hypertension [yes = 1]) + (0.3434 × known diabetes [yes = 1]) + (0.3653 × current cigarette use [yes = 1]) + (0.1849 × statin/lipid-lowering drug use [yes = 1]) + (1.0964 × baseline serum creatinine [per 1 mg/dl increase]) + (0.5403 × known RA [yes = 1]) – (0.0742 ×

Table 3. Multivariable HRs from final multivariable-adjusted models predicting primary outcome*

	Derivation cohort, HR (95% CI)	Full cohort, HR (95% CI)
Age, years	1.03 (1.02–1.04)	1.03 (1.02–1.04)
Male sex	1.31 (1.02–1.68)	1.31 (1.07–1.59)
Tobacco use	1.53 (1.17–1.99)	1.44 (1.17–1.77)
History of diabetes mellitus	1.49 (1.21–1.82)	1.41 (1.20–1.66)
History of hypertension	1.33 (1.01–1.76)	1.25 (1.01–1.56)
Prior CV event	2.75 (2.22–3.39)	2.31 (1.95–2.72)
Serum creatinine, mg/dl	2.54 (1.61–4.01)	2.99 (2.09–4.28)
Hematocrit level <43%†	0.92 (0.89–0.96)	0.93 (0.90–0.96)
Hematocrit level ≥43%†	1.06 (1.003–1.12)	1.04 (1.001–1.09)
Use of statins	1.33 (1.06–1.65)	1.20 (1.01–1.43)
RA	1.70 (1.29–2.23)	1.72 (1.37–2.15)

* All variables from Table 1 were tested in age- and sex-adjusted models. Those with hazard ratios (HRs) of >1.2 or *P* values of <0.001 were tested in adjusted models. We then removed variables with 95% confidence intervals (95% CIs) that included 1.00. CV = cardiovascular; RA = rheumatoid arthritis.
† Hematocrit was tested using linear splines with a knot set at a hematocrit level of 43%.

Table 4. Risk of major toxicity among NSAID users and component outcomes by risk category*

	No. (%)	Major NSAID toxicity, rate (95% CI)	MACE, rate (95% CI)	CSGIE, rate (95% CI)	Acute kidney injury, rate (95% CI)†	Death, rate (95% CI)‡
Low risk (<1%)	1,080 (4.6)	0.40 (0.05–0.76)	0.24 (0.0–0.52)	0.16 (0.0–0.38)	0	0.16 (0.0–0.38)
Intermediate risk (1–4%)	16,273 (68.6)	1.69 (1.49–1.89)	0.99 (0.84–1.14)	0.48 (0.38–0.59)	0.15 (0.09–0.21)	0.24 (0.16–0.31)
High risk (>4%)	6,382 (26.9)	5.56 (4.98–6.14)	3.48 (3.01–3.95)	0.84 (0.61–1.08)	1.06 (0.80–1.33)	0.83 (0.60–1.06)

* Rates are per 100 patients. NSAID = nonsteroidal antiinflammatory drug; 95% CI = 95% confidence interval; MACE = major adverse cardiovascular event; CSGIE = clinically significant gastrointestinal event.
† Acute kidney injury was defined as development of renal insufficiency or renal failure, including any of the following: serum creatinine level of ≥2.0 mg/dl and increase of ≥0.7 mg/dl from baseline, hospitalization for acute renal failure with a doubling of the baseline serum creatinine level or hyperkalemia with ≥50% elevation in serum creatinine, or initiation of dialysis.
‡ Includes all causes.

[the lesser of hematocrit level or 43]) + (0.0433 × [the greater of 0 or hematocrit level – 43]).

We then created the 3 risk groups—low risk (<1%), intermediate risk (1–4%), and high risk (>4%)—based on the predicted 1-year risk probabilities. The 1-year risk probability of major toxicity among NSAID users can be predicted with the following formula:

$$100\% \times \left(1 - 0.96855^{e^{\text{risk score}-1.05501}}\right)$$

Table 4 illustrates the number of patients in the trial cohort in each risk category and the 1-year risk of the primary outcome measure, as well as each major toxicity. In the total population with complete data (n = 23,735), 1,080 patients (4.6%) had a predicted 1-year risk of <1%, 16,273 (68.6%) had a predicted risk of 1–4%, and 6,382 (26.9%) had a predicted risk of >4%. The Kaplan-Meier survival curves for the 3 categories of risk diverged early during follow-up (Figure 1), and the log rank tests showed significant differences ($P < 0.001$).

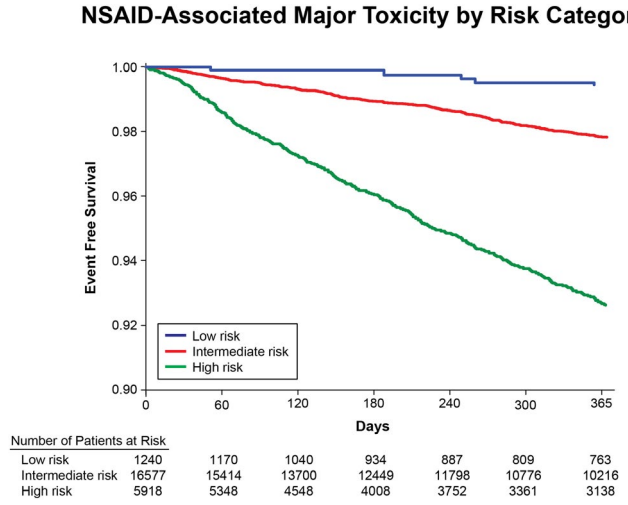

Figure 1. Kaplan-Meier survival curve illustrating the event-free survival of subjects in each of the 3 risk score categories for major toxicity among nonsteroidal antiinflammatory drug (NSAID) users. $P < 0.001$ in all log rank tests.

In subgroup analyses of specific NSAID exposure, we examined the predicted and observed risk of major toxicity among NSAID users and found strong calibration between them in the following subgroups: calibration slope was 1.0831 for celecoxib, 0.9625 for ibuprofen, and 0.9753 for naproxen (see Supplementary Figure 2, http://onlinelibrary.wiley.com/doi/10.1002/art.40870/abstract). Additionally, a separate subgroup analysis examined the performance of the risk score among patients receiving typical starting dosages versus those receiving higher dosages of the 3 NSAIDs. The risk score had strong calibration in these subgroups: calibration slope was 1.0615 for typical dosages and 0.9077 for higher dosages (Supplementary Figure 3, http://onlinelibrary.wiley.com/doi/10.1002/art.40870/abstract).

DISCUSSION

We used a large cohort of well-phenotyped patients to derive and validate a risk score for major toxicity among NSAID users. All patients had OA or RA and were enrolled in a randomized controlled trial comparing the safety of celecoxib, naproxen, and ibuprofen; they were offered concomitant esomeprazole, a proton-pump inhibitor. Adjusted models tested a broad range of potential risk factors, and 11 easily assessed variables were found to be significant in multivariable models. Calibration and discrimination statistics were moderately strong in both the derivation and the validation cohorts, and a risk score accurately predicted major toxicity among NSAID users at 1 year. The risk score defined 3 groups of patients: 4.6% of patients were at low risk (<1%), 68.6% at intermediate risk (1–4%), and 26.9% at high risk (>4%).

External validation will be important, but the split-sampling and large cohort in this study suggest that the risk score is likely to be useful in stratifying patients. Testing the risk score in populations that are not receiving proton-pump inhibitors, those with a different distribution of risk factors (i.e., fewer CV risk factors), and patients with different underlying causes of chronic pain will help determine the external validity and generalizability of the risk score. There were some slight differences between the derivation

and validation cohorts due to a midtrial protocol amendment that attempted to increase the CV event rates (Table 1). In the current analyses, all patients were receiving an NSAID, so the incremental risk associated with the use of these drugs cannot be estimated compared to nonuse.

Further analysis of the percentage of patients found to be in each of the 3 risk groups (low, intermediate, and high) can provide insight into the safety of NSAID use in adults with chronic painful conditions. In the study cohort, a relatively small group (4.6%) had a low risk (<1%) of major toxicity within 1 year, and a fairly large group (26.9%) had a high risk (>4%). We did not choose the 3 risk thresholds using formal decision analysis but rather through clinician input about what level of risk would impact decision-making. A 1-year risk of >4% translates into a number needed to harm of ≤25; this level of risk was considered to be too high to suggest taking an NSAID. Conversely, a risk of <1% translates into a number needed to harm of ≥100; this level of risk was felt to be acceptable for almost all patients. A risk between 1% and 4% was considered to be an intermediate risk, for which some providers and patients might choose that the patient take an NSAID and others not. This middle group contained the largest portion of patients in our study cohort. Future work should attempt to further break down this intermediate group, as it is likely that many patients with a 1-year risk of <2% would be willing to use an NSAID if it provided enough analgesia. We give examples of patients in all 3 risk categories in Supplementary Table 2 and provide a web calculator in Supplementary Figure 4 (http://onlinelibrary.wiley.com/doi/10.1002/art.40870/abstract).

There are several notable strengths of these analyses. First, the study data came from a prospective randomized trial with well-defined baseline characteristics, standardized adjudication of outcomes, <1% missing data, and >90% follow-up for outcomes. Additionally, using data from a randomized trial removes any confounding by indication and allowed us to develop the risk score using an equal distribution of 3 commonly prescribed NSAIDs. Second, we followed methodologic standards for risk score analyses (5). Third, the risk score performed well in the total group as well as in the subgroups divided by NSAID type. This suggests that the choice of NSAID does not substantially modify risks across the risk factors assessed. Fourth, the trial population was very large and treated with typical dosages of NSAIDs. This improves the generalizability of the risk score. Finally, the risk score variables are all easily obtained, improving the applicability of the score.

Limitations of the current analysis include its post hoc nature and the lack of an external validation cohort. The PRECISION trial population included people ages 18–100 years. Since there are few people at either extreme, estimates are unstable for the very young or the very old in this range. Another potential concern is the relatively poor long-term adherence to NSAID treatment observed in the PRECISION trial, which was the rationale for limiting the current analyses to 1 year. At 1 year, 59% of patients were still taking the study NSAID. We acknowledge that the risks cannot be directly estimated at 1 year for all patients. In the trial, as in typical practice, many patients who planned to regularly take NSAIDs did not maintain the treatment; this was the rationale for censoring patients before 1 year if they were not adherent with the study NSAID. Given the variation in follow-up time in our study, we assessed the ability of our model to predict risk of events at 6 months as well as at 1 year and found good calibration. Furthermore, the dosages of the 3 NSAIDs in the PRECISION trial have been criticized as nonequivalent. While these are valid criticisms of the trial, we included all 3 drugs at all dosages in the current set of analyses. We found that the risk score performed well across the 3 drugs and among patients receiving typical dosages and higher dosages. It is somewhat surprising that NSAID dosage was not an important modifier of the risk score performance. We did not test it as a potential predictor in the main analyses. While NSAID dosage has been found to be a risk factor for adverse events, after adjustment for all other variables in the risk score, it did not make a substantial difference.

In conclusion, if the risk score is found to be valid in external populations that may be more heterogeneous in characteristics, the risk of major NSAID toxicity can be predicted using simple clinical factors that should allow patients and providers to make a personalized decision regarding initiation of NSAIDs. Since NSAIDs represent one of the most common drug groups regularly taken in the US, safe use of these agents could provide important public health benefits. Moreover, data on the factors included in the risk score can be gleaned from most electronic medical records, making the dissemination of such a risk score relatively straightforward.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Solomon had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Solomon, Nissen, Paynter.
Acquisition of data. Nissen, Husni.
Analysis and interpretation of data. Solomon, Shao, Wolski, Paynter.

ROLE OF THE STUDY SPONSOR

Pfizer funded the PRECISION trial and supported the effort of several coauthors but had no role in the current manuscript. They were given the opportunity to review and comment on the final manuscript before submission. Publication of this article was not contingent upon approval by Pfizer.

REFERENCES

1. IMS Health. National prescription audit: oral NSAID market. 2013.
2. Wolfe MM, Lichtenstein DR, Singh G. Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. N Engl J Med 1999;340:1888–99.
3. Straube S, Tramèr MR, Moore RA, Derry S, McQuay HJ. Mortality with upper gastrointestinal bleeding and perforation: effects of time and NSAID use. BMC Gastroenterol 2009;9:41.

4. Nissen SE, Yeomans ND, Solomon DH, Lüscher TF, Libby P, Husni ME, et al. Cardiovascular safety of celecoxib, naproxen, or ibuprofen for arthritis. N Engl J Med 2016;375:2519–29.
5. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:W1–73.
6. Lloyd-Jones DM, Wilson PW, Larson MG, Beiser A, Leip EP, D'Agostino RB, et al. Framingham risk score and prediction of lifetime risk for coronary heart disease. Am J Cardiol 2004;94:20–4.
7. Graham I, Atar D, Borch-Johnsen K, Boysen G, Burell G, Cifkova R, et al. European guidelines on cardiovascular disease prevention in clinical practice: full text. Fourth Joint Task Force of the European Society of Cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). Eur J Cardiovasc Prev Rehabil 2007;14 Suppl 2:S1–113.
8. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014;129 Suppl 2:S1–45.
9. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. Osteoporos Int 2008;19:385–97.
10. Fries JF. The epidemiology of NSAID gastropathy: the ARAMIS experience. J Clin Rheumatol 1998;4 Suppl:S11–6.
11. Fries JF, Williams CA, Bloch DA, Michel BA. Nonsteroidal anti-inflammatory drug-associated gastropathy: incidence and risk factor models. Am J Med 1991;91:213–22.
12. Gabriel SE, Jaakkimainen L, Bombardier C. Risk for serious gastrointestinal complications related to use of nonsteroidal anti-inflammatory drugs: a meta-analysis. Ann Intern Med 1991;115:787–96.
13. Wolfe F. Which HAQ is best? A comparison of the HAQ, MHAQ and RA-HAQ, a difficult 8 item HAQ (DHAQ), and a rescored 20 item HAQ (HAQ20): analyses in 2,491 rheumatoid arthritis patients following leflunomide initiation. J Rheumatol 2001;28:982–9.
14. Cook N, Paynter N, Demler O. Risk prediction modeling: division of preventive medicine. 2018. URL: http://ncook.bwh.harvard.edu/sas-macros.html.