



Universiteit
Leiden
The Netherlands

Attrition and representativeness in development and validation of online symptom checkers-a case study on the Rheumatic? Questionnaire

Zegers, F.D.; Qin, L.; Selani, D.; Gomon, G.; Maarseveen, T.; Glas, K.; ... ; Knevel, R.

Citation

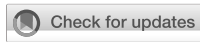
Zegers, F. D., Qin, L., Selani, D., Gomon, G., Maarseveen, T., Glas, K., ... Knevel, R. (2026). Attrition and representativeness in development and validation of online symptom checkers-a case study on the Rheumatic? Questionnaire. *Frontiers In Artificial Intelligence*, 9. doi:10.3389/frai.2026.1815241

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4304856>

Note: To cite this publication please use the final published version (if applicable).



OPEN ACCESS

EDITED BY

Vinod Ravindran,
Centre for Rheumatology, India

REVIEWED BY

Are Hugo Pripp,
Oslo University Hospital, Norway
Himanshu Pathak,
Tricolour Hospitals, India

*CORRESPONDENCE

Floor Dijkstra Zegers
✉ f.d.zegers@lumc.nl

RECEIVED 22 February 2026

REVISED 02 April 2026

ACCEPTED 09 April 2026

PUBLISHED 08 May 2026

CITATION

Zegers FD, Qin L, Selani D, Gomon G, Maarseveen T, Glas K, van Tubergen A, Ruiterman YG, Reinders M, van den Akker E, de Jong CB, Klareskog L, Axnäs B, Bos R, le Cessie S and Knevel R (2026) Attrition and representativeness in development and validation of online symptom checkers—a case study on the *Rheumatic? Questionnaire*. *Front. Artif. Intell.* 9:1815241. doi: 10.3389/frai.2026.1815241

COPYRIGHT

© 2026 Zegers, Qin, Selani, Gomon, Maarseveen, Glas, van Tubergen, Ruiterman, Reinders, van den Akker, de Jong, Klareskog, Axnäs, Bos, le Cessie and Knevel. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Attrition and representativeness in development and validation of online symptom checkers—a case study on the *Rheumatic? Questionnaire*

Floor Dijkstra Zegers^{1,2*}, Ling Qin², Daniyal Selani³, Georgy Gomon^{1,2}, Tjardo Maarseveen², Kasper Glas⁴, Astrid van Tubergen^{5,6}, Yvonne Goekoop Ruiterman⁷, Marcel Reinders^{3,8}, Erik van den Akker^{3,8}, Corne Baatenburg de Jong⁹, Lars Klareskog¹⁰, Barbara Axnäs¹¹, Reinhard Bos^{12,13}, Saskia le Cessie^{1,8} and Rachel Knevel²

¹Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands,

²Department of Rheumatology, Leiden University Medical Center, Leiden, Netherlands, ³Department of Intelligent Systems, Pattern Recognition and Bioinformatics, Delft University of Technology, Delft, Netherlands, ⁴Reumazorg Zuid-West Nederland, Roosendaal, Netherlands, ⁵Department of Rheumatology, Maastricht University Medical Centre+, Maastricht, Netherlands, ⁶Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, Netherlands, ⁷Department of Clinical Epidemiology, Haga Hospital, The Hague, Netherlands, ⁸Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands, ⁹Research, Innovation and Impact Funding, ReumaNederland (Dutch Arthritis Society), Amsterdam, Netherlands, ¹⁰Division of Rheumatology, Department of Medicine, Karolinska Institutet, Stockholm, Sweden, ¹¹Elsa Science, Stockholm, Sweden, ¹²Department of Rheumatology, Medical Center Leeuwarden, Leeuwarden, Netherlands, ¹³Department of Rheumatology, Nij Smellinghe Hospital, Drachten, Netherlands

Background: Online symptom checkers are often developed and validated on data subject to self-selection and selective attrition, potentially introducing biases in prediction models.

Objectives: To assess recruitment, selection, and attrition patterns in a large Dutch online symptom checker for musculoskeletal complaints and to evaluate potential biases by comparing participant characteristics across recruitment sources and with external target populations.

Methods: Using data from the online Dutch *Rheumatic? Questionnaire* on musculoskeletal complaints, we compared baseline characteristics and key self-reported symptoms between responders to the follow-up survey and non-responders. The survey responders were furthermore compared according to source of recruitment to the questionnaire, i.e., via primary care clinics, secondary care clinics, or via different online sources. Sex, age and BMI distributions from the total study group were compared to external data of potential target populations of primary and secondary care patients within the Netherlands.

Results: The total study group of answers to the questionnaire comprised 31,457 responders, of which 50% ($n = 15,591$) responded to the follow-up survey. Study participants were predominantly female (76%), middle-aged (one-third 50–60 years), never-smokers (66%), and overweight. While participants recruited through healthcare settings resembled target populations, follow-up survey responders were older, had more rheumatic diagnoses (49% vs. 32%), and reported more symptoms than non-responders. Participant characteristics varied by recruitment source, with social media attracting younger females

while healthcare routes reached more diverse populations with varying symptom presentations.

Conclusion: Patterns of recruitment and attrition produced differences in participant characteristics. Healthcare-based recruitment yielded participants resembling intended target populations, and follow-up survey responders differed on some points from nonresponders. Awareness of these selection processes is essential when using real-world symptom checker data for model development.

KEYWORDS

digital decision support system, generalizability, health informatics, musculoskeletal complaints, prediction model, real-world data

Introduction

Online symptom checkers (OSCs) and digital triage tools promise the possibility of optimizing the diagnostic process, as well as aiding health professionals in decision making, therefore aiding the promotion of health equity (Abramoff et al., 2023). For an OSC to be of value, it is crucial that it performs well on the intended target population. Many OSCs rely solely on textbook knowledge or are validated only using clinical vignettes or physician review of patient records (Wallace et al., 2022; El-Osta et al., 2022; Fraser et al., 2018), limiting their validity. Current recommendations promote that OSCs should be developed and validated using real-world data from broadly representative populations (Wallace et al., 2022; Sperrin et al., 2022).

However, using real-world data introduces methodological challenges. The PROBAST (Prediction model Risk Of Bias ASsessment Tool) guidelines describes several domains of risk of bias when using real-world data for model development (Wolff et al., 2019). Particularly, they highlight selective participation and incomplete follow-up. If certain groups such as healthier users or individuals with milder symptoms, are less likely to participate or complete follow-up, systematic bias may arise, leading to reduced generalizability of prediction models to the intended users (Wolff et al., 2019; Lasko et al., 2024; Bethlehem, 2010; Gustavson et al., 2012; de Leeuw and Lugtig, 2015; Keiding and Louis, 2016). Selective dropout may further distort associations between predictors and outcomes, particularly when outcomes depend on follow-up surveys. Therefore, evaluating recruitment sources, participant characteristics, and follow-up completeness is essential to understand potential selection bias when real-world data are used for OSC development.

Although potential problems with selective participation, such as the risk of collider bias, have been discussed (Munafo et al., 2018), few studies so far have described the selection processes in real-world online recruitment for OSCs.

The European SPIDeRR project (Stratification of Patients using advanced Integrative modeling of Data Routinely acquired for diagnosing rheumatic complaints) offers a relevant test case. The project aims to develop an OSC specifically for people with musculoskeletal complaints (MSCs). Musculoskeletal symptoms are highly frequent, comprising 22% of the consultations of primary care physicians in the Netherlands (der van Linden et al., 2004; Finckh, 2009; Raciborski et al., 2017; Jain et al., 2023; Knitza et al., 2021). Despite the benefits of early diagnosis, diagnostic delays remain common (Raciborski et al., 2017), prompting many individuals to seek online explanations for their symptoms (Jain et al., 2023; Knitza et al., 2021; Knevel et al., 2022).

SPIDeRR's OSC, *Rheumatic?* (Knevel et al., 2022) is a questionnaire collecting information on symptoms, past diagnoses and basic characteristics (Lundberg et al., 2023). The OSC aims to provide accurate disease probabilities for different settings: (1) people online with MSCs, (2) general practitioners (GPs) using the OSC for triage and referral, and (3) for rheumatologists to optimize the diagnostic process. Participants are followed for 1 year to collect information on subsequent referral, diagnosis and medication.

In this study, we assessed recruitment, selection, and attrition patterns among users of the *Rheumatic?* Questionnaire. We compared participant characteristics across recruitment sources and with external data from Dutch primary and secondary care populations with MSCs. In addition, we examined selective follow-up by comparing baseline-only participants with those who completed the first follow-up survey.

Methods

Study design and setting

Rheumatic? (Elsa Science AB, 2024) is an online questionnaire about current symptoms, past diagnoses and demographic information (all questions provided in [Supplementary Table 1](#)). Participants were recruited through multiple channels: online advertisements, the website and newsletters of the Dutch Arthritis Society (ReumaNederland), primary care (GP) clinics, and secondary care rheumatology outpatient clinics. Participants consented to be contacted for follow-up. Four follow-up surveys were sent via e-mail. The first immediately after the baseline questionnaire (included questions about source of recruitment, user-experience and further diagnoses), and the last three after 3, 6 and 12 months, including questions regarding subsequent rheumatologic diagnoses, medication usage, GP visits and referrals.

Data processing

We included Dutch adults who completed *Rheumatic?* Between 24 July 2021 and 13 July 2023 and provided informed consent. Responses were excluded if demographic information was missing or if email addresses were missing, duplicated, or invalid, ensuring that follow-up questionnaires could be delivered and linked. For the first follow-up survey, only fully completed surveys submitted within 7 days of baseline were included.

Variables

For the present study we considered age, sex, height, weight, smoking and alcohol habits, rheumatic diagnosis within near family and key symptoms (pain, morning stiffness, all day stiffness, exhaustion, reduced endurance or none). Age was classified according to <40, 40–50, 50–60 and 60+ years. Body mass index (BMI) was derived from midpoint values of weight and height categories (5 kg and 5 cm intervals) and grouped as <18.5, 18–24.9, 25–30 and 30+ (Weir and Jan, 2024). Family diagnosis was defined as reporting a rheumatic-related diagnosis within close family (yes/no/do not know). Previous diagnosis was defined as reporting any previously given rheumatic-related diagnosis at baseline or in the follow-up survey (yes/no/do not know). Recruitment sources (via a rheumatology clinic, GP, the ReumaNederland website/newsletters, other online sources (with a free text option) or other sources) and further rheumatic diagnoses were solely indicated in the follow-up survey.

Website analytics

Daily inclusion counts were compared with daily numbers of Dutch new visitors to the website (Elsa Science AB, 2024) hosting the questionnaire between July 2021 and December 2023. All website analytics were summarized using Google Analytics (Google Analytics, 2024).

Expected target populations

To assess representativeness, we collected data from the same time-period (24 July 2021 to 13 July 2023) from (i) all GP episodes with MSCs-related ICD-codes (L01–L20, L29, L83–L94, L98, L99, T92, thus excluding MSC consultations with fractures, infections and innate diseases) (Hofmans-Okkes and Lamberts, 1996; Ardesch et al., 2023), (ii) regional rheumatology clinics in the South West Netherlands (RZWN), and (iii) the academic rheumatology clinic of Leiden University Medical Centre (LUMC). Age, sex and BMI (when available) were compared between the study population and these potential target populations.

Statistical analysis

Contingency tables summarized baseline characteristics of all participants and compared responders vs. nonresponders to the first follow-up survey. Differences in percentages were presented with 95% confidence intervals (CIs) and visualized in a forest plot.

Recruitment source differences were assessed using contingency tables stratified by reported source of recruitment coded as GP, rheumatology clinic, other sources in general, and online sources. The latter was subdivided based on free text answers into the ReumaNederland website, Facebook, Google, Instagram, or other online sources. We used only participants' answers to the first follow-up survey since this survey contained information on sources of recruitment.

Representativeness relative to target populations was assessed by comparing distributions of age, sex, and BMI between responders to the follow-up survey and external clinical populations, grouped as above. In the external data, BMI was only calculated when available.

Recruitment patterns over time were examined using frequency plots of inclusion date and Google hit-data with 7-day rolling means

and furthermore stratified by recruitment source for responders to the follow-up survey.

All data handling and analysis were performed using R version 4.1.3 (R Core Team, 2023), using ggplot2 for plots (Wickham, 2016), and gt and gtsummary for tables (Sjoberg et al., 2021; Iannone et al., 2024).

Results

Characteristics of all *Rheumatic?* Participants

After data processing 31,159 participants were included (Supplementary Figure 1). Participants were predominantly female (76%), most commonly aged 50 to 60 years (31%), never-smokers (66%) and frequently overweight. About half (51%) reported no previous rheumatic diagnosis, and 48% reported a close family member having a rheumatic diagnosis. Pain (83%) and morning stiffness (62%) were the most frequent symptoms. 29% reported swelling in one or more body parts and 1% reported having none of the key symptoms. Characteristics of participants are fully specified in Supplementary Table 2.

Differences between follow-up survey participants and non-participants

Of all participants, half ($N = 15,591$) responded to the first follow-up survey. Characteristics of follow-up responders and nonresponders are presented in Supplementary Table 2, and differences in percentage are shown in Figure 1 and Supplementary Table 2. Follow-up participation did not yield differences in sex, alcohol consumption, family history of rheumatic disease, or having a complete absence of key symptoms. However, responders were older, more frequently reported a prior rheumatic diagnosis (49% vs. 32%), were less frequently current smokers and more often had a BMI above 25. Responders more frequently reported key symptoms, except for exhaustion which was reported more frequently among nonresponders than responders.

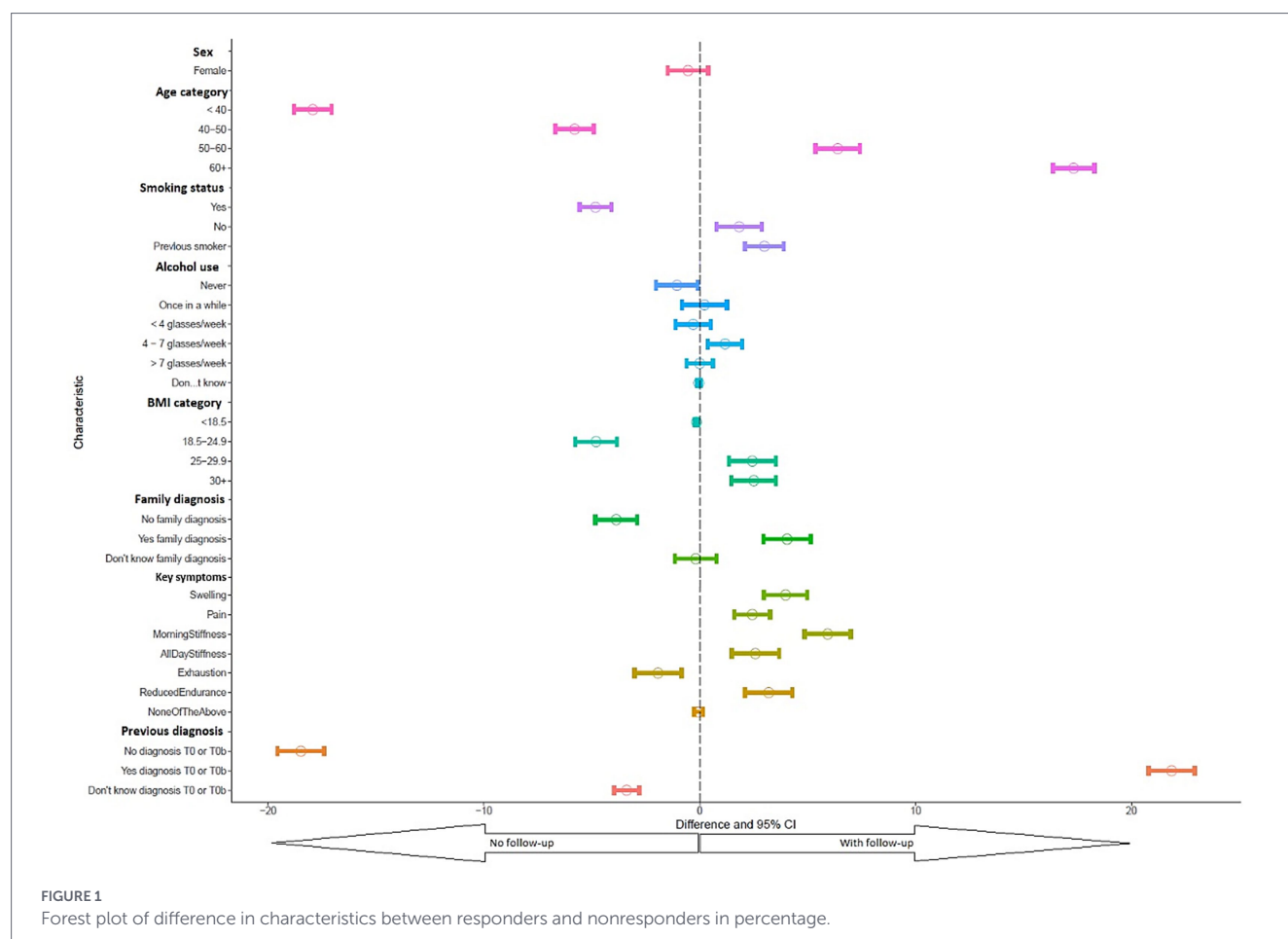
Characteristics by source of recruitment

Most of the participants with information on recruitment [12,574 (81%)] were recruited online, with the largest proportion originating from ReumaNederland communication channels, Google, and social media platforms (Table 1). Smaller proportions were recruited through secondary care rheumatology clinics, general practitioners (GPs), or other offline sources.

Social media recruitment predominantly reached younger women, with women accounting for 95 to 98% of participants recruited via Facebook and Instagram. These participants also showed slightly higher BMI values, with the highest prevalence of obesity observed among Facebook recruits.

Other online recruitment sources, including Google and disease-specific websites, reached a more heterogeneous population. Participants recruited via Google less often had a prior rheumatic diagnosis.

Healthcare-based recruitment through GPs and rheumatology clinics resulted in more balanced sex distributions and older age profiles. Participants recruited via rheumatology clinics most frequently reported swelling, while those recruited through GPs reported fewer symptoms overall (Figure 2).



Family history of rheumatic disease varied modestly across recruitment sources, and smoking patterns showed little variation.

Recruitment tendencies

Daily recruitment rates varied over time and are shown in [Figure 3](#). Distinct recruitment peaks coincided with newsletter campaigns conducted by ReumaNederland, annual World Arthritis Day and targeted social media campaigns. Source-stratified analyses highlighted the effect of paid social media campaigns on increases in recruitment, whereas organic posts on the ReumaNederland social media site had limited impact ([Supplementary Figure 2](#)). Daily recruitment closely followed traffic to the questionnaire website ([Supplementary Figure 3](#)).

Comparison with target populations in primary and secondary care

Compared with patients from Dutch secondary care rheumatology clinics, the overall *Rheumatic?* Cohort included a higher proportion of women. However, participants recruited via secondary care closely resembled secondary care patients with respect to sex. Patients from the academic rheumatology clinic were younger and were more present in the lower BMI groups. However, participants from the regional rheumatology clinic resembled *Rheumatic?* Rheumatology recruits more closely ([Table 2](#)).

In comparison with primary care data from GPs, women were overrepresented in the overall *Rheumatic?* Cohort. This imbalance was not observed among participants recruited through GPs, whose sex distribution closely matched that of the primary care population. The overall *Rheumatic?* Cohort held relatively more middle-aged participants (50–60 years), and those recruited via GPs tended to be older than individuals in the external GP data. BMI distributions were comparable between all groups ([Table 2](#)).

Discussion

In this cohort study of more than 31,000 individuals with musculoskeletal symptoms using the *Rheumatic?* Online symptom checker (OSC), we identified distinct recruitment, selection, and attrition patterns with direct implications for representativeness and potential bias in real-world OSC data. Recruitment through primary and secondary care settings yielded participant characteristics closely resembling their intended target populations, whereas online recruitment—particularly via social media—resulted in younger and predominantly female participants. In addition, follow-up survey responders differed systematically from nonresponders, indicating systematic attrition.

Our findings are consistent with prior studies of non-disease-specific OSCs, which reported a predominance of female users (62–85%) and younger mean ages ([Kopka et al., 2023](#); [Winn et al., 2019](#); [Arellano Carmona et al., 2022](#); [Morse et al., 2020](#)). Pain was also the

TABLE 1 Characteristics according to source of recruitment for participants responding to the follow-up survey.

Characteristic	Facebook, N = 929 (6.0%)	Instagram, N = 289 (1.9%)	Google, N = 1,266 (8.1%)	ReumaNL ^b , N = 9,589 (62%)	Other online, N = 501 (3.2%)	GP, N = 98 (0.6%)	Rheumatology, N = 1,267 (8.1%)	Other, N = 1,652 (11%)
Sex								
Female	881 (95%)	282 (98%)	933 (74%)	7,181 (75%)	396 (79%)	54 (55%)	813 (64%)	1,304 (79%)
Age category								
<40	59 (6.4%)	45 (16%)	255 (20%)	1,184 (12%)	32 (6.4%)	13 (13%)	134 (11%)	141 (8.5%)
40–50	200 (22%)	65 (22%)	290 (23%)	1,731 (18%)	48 (9.6%)	9 (9.2%)	172 (14%)	238 (14%)
50–60	541 (58%)	139 (48%)	380 (30%)	3,162 (33%)	160 (32%)	23 (23%)	378 (30%)	591 (36%)
60+	129 (14%)	40 (14%)	341 (27%)	3,512 (37%)	261 (52%)	53 (54%)	583 (46%)	682 (41%)
Smoking status								
Yes	103 (11%)	19 (6.6%)	157 (12%)	956 (10.0%)	37 (7.4%)	14 (14%)	175 (14%)	175 (11%)
No	621 (67%)	217 (75%)	808 (64%)	6,488 (68%)	343 (68%)	62 (63%)	791 (62%)	1,084 (66%)
Previous smoker	205 (22%)	53 (18%)	301 (24%)	2,145 (22%)	121 (24%)	22 (22%)	301 (24%)	393 (24%)
Alcohol use								
Never	245 (26%)	74 (26%)	294 (23%)	2,365 (25%)	120 (24%)	19 (19%)	331 (26%)	430 (26%)
Once in a while	376 (40%)	113 (39%)	458 (36%)	3,173 (33%)	143 (29%)	41 (42%)	414 (33%)	602 (36%)
<4 glasses/week	137 (15%)	55 (19%)	201 (16%)	1,604 (17%)	98 (20%)	13 (13%)	193 (15%)	249 (15%)
4–7 glasses/week	114 (12%)	41 (14%)	186 (15%)	1,604 (17%)	86 (17%)	17 (17%)	233 (18%)	241 (15%)
>7 glasses/week	56 (6.0%)	6 (2.1%)	124 (9.8%)	833 (8.7%)	53 (11%)	7 (7.1%)	89 (7.0%)	126 (7.6%)
Do not know	1 (0.1%)	0 (0%)	3 (0.2%)	10 (0.1%)	1 (0.2%)	1 (1.0%)	7 (0.6%)	4 (0.2%)
BMI category								
<18.5	0 (0%)	1 (0.3%)	1 (<0.1%)	4 (<0.1%)	0 (0%)	0 (0%)	0 (0%)	2 (0.1%)
18.5–24.9	130 (14%)	66 (23%)	349 (28%)	2,343 (24%)	122 (24%)	22 (22%)	273 (22%)	356 (22%)
25–29.9	396 (43%)	124 (43%)	509 (40%)	4,142 (43%)	224 (45%)	47 (48%)	541 (43%)	696 (42%)
30+	403 (43%)	98 (34%)	407 (32%)	3,100 (32%)	155 (31%)	29 (30%)	453 (36%)	598 (36%)
Family diagnosis								
No	250 (27%)	75 (26%)	338 (27%)	2,351 (25%)	113 (23%)	21 (21%)	314 (25%)	375 (23%)
Yes	486 (52%)	166 (57%)	587 (46%)	4,826 (50%)	273 (54%)	45 (46%)	625 (49%)	843 (51%)
Do not know	193 (21%)	48 (17%)	341 (27%)	2,412 (25%)	115 (23%)	32 (33%)	328 (26%)	434 (26%)
Previous diagnosis^a								
No	367 (40%)	137 (47%)	694 (55%)	3,900 (41%)	175 (35%)	34 (35%)	395 (31%)	523 (32%)
Yes	528 (57%)	138 (48%)	474 (37%)	5,174 (54%)	301 (60%)	55 (56%)	764 (60%)	1,031 (62%)
Do not know	34 (3.7%)	14 (4.8%)	98 (7.7%)	515 (5.4%)	25 (5.0%)	9 (9.2%)	108 (8.5%)	98 (5.9%)

^aSelf-reported rheumatic-related diagnosis at baseline or in the follow-up survey.

^bReumaNL, ReumaNederland.

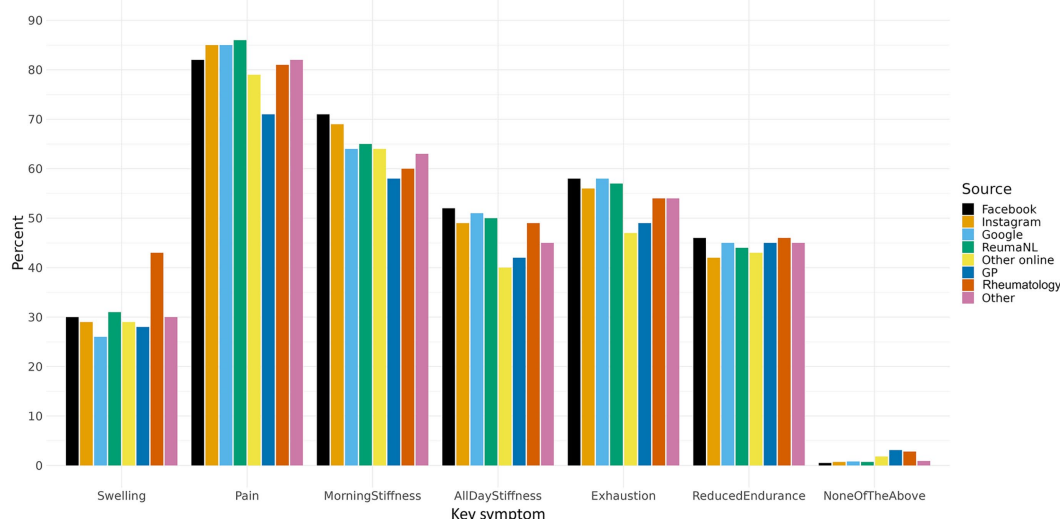


FIGURE 2
Barplot of reported key symptoms according to recruitment source.

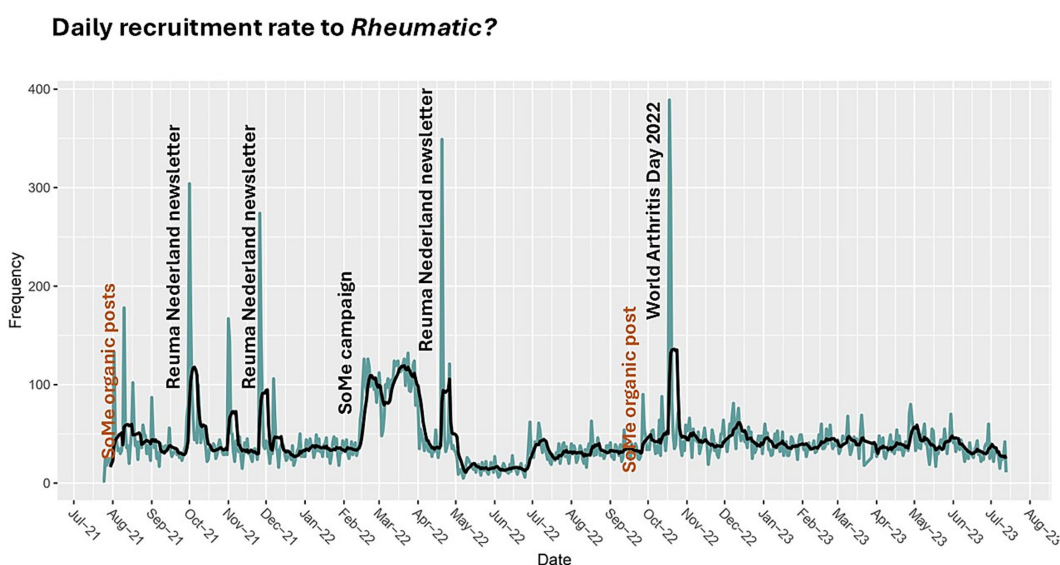


FIGURE 3
Daily recruitment rate in the study.

most frequently reported symptom in those studies, mirroring our observations. Studies on OSCs specifically targeting musculoskeletal complaints remains limited. A small German study evaluating musculoskeletal-focused OSCs and an Australian primary care study reported sex and age distributions broadly comparable to our findings, although sample sizes were substantially smaller (Knitz et al., 2021; Haas et al., 2023). Together, these data suggest that online symptom checkers for musculoskeletal complaints consistently attract selective user groups, particularly when participants are recruited outside healthcare settings.

Recruitment source and representativeness

Participants recruited through healthcare settings closely matched external primary and secondary care populations with

respect to sex, age, and BMI. This alignment suggests that OSC deployment within clinical pathways may yield datasets representative of these intended users. In contrast, participants recruited online were younger and more frequently female, with notable variation between online recruitment channels. Social media based recruitment resulted in an almost exclusively female population, whereas Google-based recruitment reached a more heterogeneous group with fewer prior diagnoses. Importantly, our findings demonstrate that “online users” do not constitute a homogeneous population. Marked differences in demographic and clinical characteristics were observed across online recruitment sources. As other authors previously have suggested, these patterns likely reflect differences in platform use, content targeting, and health-seeking behavior, underscoring that recruitment strategy itself is a key determinant of dataset composition (Bol et al., 2020).

TABLE 2 Comparison of the participants responding to the follow-up survey with external data.

Reported characteristic <i>n</i> (%)	<i>Rheumatic?</i> Questionnaire			External	External	
	All <i>N</i> = 15,591	Recruited via GP, <i>N</i> = 98	Recruited via rheumatology clinics, <i>N</i> = 1,267	Primary care ELAN ^{a,b} , <i>N</i> = 303,245	Rheumatology clinics RZWN ^{a,c} , <i>N</i> = 14,195	LUMC ^{a,d} , <i>N</i> = 4,040
Sex						
Female	11,844 (76%)	54 (55%)	813 (64%)	177,077 (59%)	9,567 (67%)	2,623 (65%)
Missing	0	0	0	4,540	0	0
Age						
<40	1,863 (12%)	13 (13%)	134 (11%)	52,238 (18%)	2,693 (19%)	992 (25%)
40–50	2,753 (18%)	9 (9.2%)	172 (14%)	52,957 (18%)	2,333 (16%)	691 (17%)
50–60	5,374 (34%)	23 (23%)	378 (30%)	67,071 (22%)	3,457 (24%)	888 (22%)
60+	5,601 (36%)	53 (54%)	583 (46%)	126,155 (42%)	5,712 (40%)	1,469 (36%)
Missing	0	0	0	4,824	0	0
BMI						
<18.5	35 (0.2%)	0 (0%)	2 (0.2%)	1,590 (1%)	188 (1.5%)	26 (2%)
18.5–24.9	3,771 (24%)	24 (24%)	285 (22%)	43,625 (25%)	4,155 (33%)	465 (40%)
25–29.9	6,542 (42%)	45 (46%)	527 (42%)	70,727 (40%)	4,048 (32%)	394 (34%)
30+	5,243 (34%)	29 (30%)	453 (36%)	59,608 (34%)	4,290 (34%)	277 (24%)
Missing	0	0	0	127,695	1,514	2,878

^aPercentages calculated among non-missing.
^bELAN, Extramural LUMC Academic Network.
^cRZWN, Reumazorg Zuid West Nederland.
^dLUMC, Leiden University Medical Center.

Selective attrition and follow-up

Follow-up survey responders differed from nonresponders in age, symptom burden, smoking status, and prevalence of prior rheumatic diagnoses. A follow-up response rate of 50% is comparable to other longitudinal studies, including surveys in rheumatology and other domains (Garcia et al., 2014; Wu et al., 2022). The higher rate of follow-up participation among older participants and nonsmokers aligns with previous population-based studies (Okpara et al., 2023; Canivet et al., 2021; Bamer et al., 2020). However, the greater symptom burden and higher prevalence of prior diagnoses among responders suggest that disease severity and health awareness influenced continued participation.

Implications for prediction model development

Selective attrition complicates the interpretation and generalizability of analyses relying on follow-up data. If associations between predictors and outcomes differ between responders and nonresponders, prediction models derived from follow-up data may be biased. Multiple imputation and sensitivity analyses under alternative missingness mechanisms may be considered to assess robustness (Sterne et al., 2009; van Buuren, 2018). Where feasible, linkage to external data sources may help validate outcomes and improve handling of missing data. In the *Rheumatic?* Study, planned linkage with

electronic health records from primary and secondary care may address these limitations.

Before developing prediction models using OSC data, the intended use and target population should be clearly defined. For *Rheumatic?*, potential applications range from a general online self-assessment tool to decision support in primary or secondary care. Each use case implies a different target population and, consequently, different requirements for representativeness.

When discrepancies between the study population and target population exist, statistical approaches such as reweighting or poststratification may improve transportability (Wolff et al., 2019; Keiding and Louis, 2016; Petersen et al., 2024; Zhang et al., 2014). Additionally, when recruitment sources are strongly associated with participant characteristics, incorporating recruitment source as a predictor or exploring interactions with other predictors may be warranted. Separate validation of prediction models across recruitment sources may further enhance robustness. More research is needed to compare different weighting and calibration methods in the case of making prediction models with multiple target groups from a heterogeneous population.

Handling missing follow-up data

Missing follow-up data introduces additional methodological challenges. When missingness is associated with participant characteristics, standard assumptions such as missing at random must be carefully evaluated. Multiple imputation and sensitivity

analyses under alternative missingness mechanisms may be considered to assess robustness (Sterne et al., 2009; van Buuren, 2018; Biering et al., 2015). Where feasible, linkage to external data sources may help validate outcomes and improve handling of missing data. In the *Rheumatic?* Study, planned linkage with electronic health records from primary and secondary care may address these limitations.

Strengths and limitations

This study provides one of the first large-scale, empirical descriptions of recruitment and participation patterns in an OSC targeting musculoskeletal complaints. The large sample size enabled detailed stratified analyses across recruitment sources and follow-up status. Although formal external reference data for the purely online target population were unavailable, the observed recruitment and attrition patterns reflect well-described mechanisms of self-selection and differential participation in digital health research. These mechanisms are unlikely to be specific to musculoskeletal complaints and may be relevant to a broad range of online symptom checkers and digital triage tools. This study is limited by incompleteness of external reference data: BMI was incomplete, and we lacked measures of socioeconomic status and digital literacy which could have provided extra insight in the comparison. Furthermore, recruitments peaks tied to specific campaigns may have altered the sample composition, limiting inferences about routine use of the symptom checker.

Conclusion

Our study underlines the challenges that arise when recruiting real-world participants online. Recruitment and attrition processes play a central role in shaping the composition of real-world online symptom checker datasets. In this study, we observed that patterns of recruitment and attrition created differences in participant characteristics. However, we also observed that recruitment via general practitioners and rheumatology clinics yielded participant samples resembling their intended clinical target populations. These patterns reflect general mechanisms of self-selection and differential participation in digital health research and underscore the importance of explicitly evaluating recruitment and attrition when developing and validating prediction models using real-world OSC data.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. All scripts used are publicly available at: https://github.com/fdzegers/Selection_and_attrition.

Ethics statement

The studies involving humans were approved by the LUMC Ethical Committee (MREC LDD). The studies were conducted in accordance

with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

FZ: Writing – original draft, Formal analysis, Investigation, Resources, Data curation, Software, Writing – review & editing, Methodology, Project administration, Conceptualization, Visualization, Validation. LQ: Writing – review & editing, Data curation. DS: Investigation, Writing – review & editing, Methodology, Data curation. GG: Methodology, Writing – review & editing, Data curation. TM: Data curation, Writing – review & editing, Investigation, Methodology. KG: Writing – review & editing, Data curation. AT: Writing – review & editing, Investigation. YR: Data curation, Writing – review & editing, Investigation. MR: Investigation, Methodology, Writing – review & editing. EA: Methodology, Investigation, Writing – review & editing. CJ: Investigation, Writing – review & editing. LK: Writing – review & editing. BA: Writing – review & editing, Investigation, Formal analysis, Data curation. RB: Writing – review & editing, Investigation. SC: Methodology, Writing – review & editing, Investigation, Supervision, Writing – original draft, Conceptualization, Data curation, Visualization, Formal analysis, Project administration, Validation. RK: Project administration, Funding acquisition, Visualization, Resources, Formal analysis, Validation, Data curation, Writing – review & editing, Methodology, Supervision, Conceptualization, Writing – original draft, Investigation, Software.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was supported by Horizon-EU within the SPIDERR-Project (Activity No. 101080711), and the DigiPrevent Programme (funded in April 2022) from EIT Health. The project has further received funding by the Dutch Organization for Health Research and Development (ZonMw) via the Klinische Fellow No. 40-00703-97-19069, and Open Competitie, No. 09120012110075. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Conflict of interest

AT: Galapagos, Johnson and Johnson, Novartis and UCB. BA: ELSA Science. RB: Galapagos, Sanofi and consulting fees from Abbvie, Galapagos, Janssen, Pfizer, UCB.

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frai.2026.1815241/full#supplementary-material>

References

- Abramoff, M. D., Tarver, M. E., Loyo-Berrios, N., Trujillo, S., Char, D., Obermeyer, Z., et al. (2023). Considerations for addressing bias in artificial intelligence for health equity. *npj Digit. Med.* 6:170. doi: 10.1038/s41746-023-00913-9
- Ardesch, F. H., Meulendijk, M. C., Kist, J. M., Vos, R. C., Vos, H. M. M., de Kieffe-Jong, J. C., et al. (2023). The introduction of a data-driven population health management approach in the Netherlands since 2019: the extramural LUMC academic network data infrastructure. *Health Policy* 132:104769. doi: 10.1016/j.healthpol.2023.104769
- Arellano Carmona, K., Chittamuru, D., Kravitz, R. L., Ramondt, S., and Ramirez, A. S. (2022). Health information seeking from an intelligent web-based symptom checker: cross-sectional questionnaire study. *J. Med. Internet Res.* 24:e36322. doi: 10.2196/36322
- Bamer, A. M., McMullen, K., Gibran, N., Holavanahalli, R., Schneider, J. C., Carrougher, G. J., et al. (2020). Factors associated with attrition of adult participants in a longitudinal database: a national institute on disability, independent living, and rehabilitation research burn model system study. *J. Burn Care Res.* 41, 270–279. doi: 10.1093/jbcr/irz186
- Bethlehem, J. (2010). Selection bias in web surveys. *Int. Stat. Rev.* 78, 161–188. doi: 10.1111/j.1751-5823.2010.00112.x
- Biering, K., Hjollund, N. H., and Frydenberg, M. (2015). Using multiple imputation to deal with missing data and attrition in longitudinal studies with repeated measures of patient-reported outcomes. *Clin. Epidemiol.* 7, 91–106. doi: 10.2147/CLEP.S72247
- Bol, N., Strycharz, J., Helberger, N., van de Velde, B., and de Vreese, C. H. (2020). Vulnerability in a tracked society: combining tracking and survey data to understand who gets targeted with what content. *New Media Soc.* 22, 1996–2017. doi: 10.1177/1461444820924631
- Canivet, C., Nilsson, A., Björk, J., Moghaddassi, M., and Östergren, P. O. (2021). Assessment of selection bias due to dropouts in the follow-up of the Scania public health cohort. *Scand. J. Public Health* 49, 457–464. doi: 10.1177/1403494820919544
- de Leeuw, E. D., and Lugtig, P. (2015). "Dropouts in longitudinal surveys," in *Wiley StatsRef: Statistics Reference Online*, (Hoboken, NJ: John Wiley & Sons, Ltd.), 1–6.
- der van Linden, M., Westert, G., de Bakker, D., and Schellevis, F. (2004). *Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Klachten en aandoeningen in de bevolking en in de huisartspraktijk*. Utrecht: NIVEL.
- El-Osta, A., Webber, I., Alaa, A., Bagkeris, E., Mian, S., Sharabiani, M. T. A., et al. (2022). What is the suitability of clinical vignettes in benchmarking the performance of online symptom checkers? An audit study. *BMJ Open* 12:e053566. doi: 10.1136/bmjopen-2021-053566
- Elsa Science AB. (2024). Available online at: <https://rheumatic.elsa.science/en/>. (Accessed April 4, 2025)
- Finckh, A. (2009). Early inflammatory arthritis versus rheumatoid arthritis. *Curr. Opin. Rheumatol.* 21, 118–123. doi: 10.1097/BOR.0b013e3283235ac4
- Fraser, H., Coiera, E., and Wong, D. (2018). Safety of patient-facing digital symptom checkers. *Lancet* 392, 2263–2264. doi: 10.1016/S0140-6736(18)32819-8
- Garcia, I., Portugal, C., Chu, L. H., and Kawatkar, A. A. (2014). Response rates of three modes of survey administration and survey preferences of rheumatoid arthritis patients. *Arthritis Care Res.* 66, 364–370. doi: 10.1002/acr.22125
- Google Analytics. (2024). Available online at: <http://google.com/analytics>. (Accessed March 13, 2024).
- Gustavson, K., von Soest, T., Karevold, E., and Roysamb, E. (2012). Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health* 12:918. doi: 10.1186/1471-2458-12-918
- Haas, R., Gorelik, A., Busija, L., O'Connor, D., Pearce, C., Mazza, D., et al. (2023). Prevalence and characteristics of musculoskeletal complaints in primary care: an analysis from the population level and analysis reporting (POLAR) database. *BMC Prim. Care* 24:40. doi: 10.1186/s12875-023-01976-z
- Hofmans-Okkes, I., and Lamberts, H. (1996). The international classification of primary care (ICPC): new applications in research and computer-based patient records in family practice. *Fam. Pract.* 13, 294–302. doi: 10.1093/fampra/13.3.294
- Iannone, R. C. J., Schloerke, B., Hughes, E., Lauer, A., and Seo, J. (2024). gt: Easily create presentation-ready display tables R package version 1.3.0.9000. Available online at: <https://github.com/rstudio/gt>
- Jain, A., Joseph, S., James, J., James, T. S., Kumar, K., Raza, K., et al. (2023). Delay in diagnosis of rheumatoid arthritis: reasons and trends over a decade. *Rheumatol. Int.* 43, 503–508. doi: 10.1007/s00296-022-05187-0
- Keiding, N., and Louis, T. A. (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *J. R. Stat. Soc. A* 179, 319–376. doi: 10.1111/rssa.12136
- Knevel, R., Knitza, J., Hensvold, A., Circiumaru, A., Bruce, T., Evans, S., et al. (2022). Rheumatic?—a digital diagnostic decision support tool for individuals suspecting rheumatic diseases: a Multicenter pilot validation study. *Front. Med.* 9:774945. doi: 10.3389/fmed.2022.774945
- Knitza, J., Mohn, J., Bergmann, C., Kampylafka, E., Hagen, M., Bohr, D., et al. (2021). Accuracy, patient-perceived usability, and acceptance of two symptom checkers (Ada and Rheport) in rheumatology: interim results from a randomized controlled crossover trial. *Arthritis Res. Ther.* 23:112. doi: 10.1186/s13075-021-02498-8
- Kopka, M., Scatturin, L., Napierala, H., Furstenau, D., Feufel, M. A., Balzer, E., et al. (2023). Characteristics of users and nonusers of symptom checkers in Germany: cross-sectional survey study. *J. Med. Internet Res.* 25:e46231. doi: 10.2196/46231
- Lasko, T. A., Strobl, E. V., and Stead, W. W. (2024). Why do probabilistic clinical models fail to transport between sites? *npj Digit. Med.* 7:53. doi: 10.1038/s41746-024-01037-4
- Lundberg, K., Qin, L., Aulin, C., van Spil, W. E., Maurits, M. P., and Knevel, R. (2023). Population-based user-perceived experience of Rheumatic?: a novel digital symptom-checker in rheumatology. *RMD Open* 9:e002974. doi: 10.1136/rmdopen-2022-002974
- Morse, K. E., Ostberg, N. P., Jones, V. G., and Chan, A. S. (2020). Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. *J. Med. Internet Res.* 22:e20549. doi: 10.2196/20549
- Munafo, M. R., Tilling, K., Taylor, A. E., Evans, D. M., and Davey, S. G. (2018). Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.* 47, 226–235. doi: 10.1093/ije/dyx206
- Okpara, C., Adachi, J., Papaioannou, A., Ioannidis, G., and Thabane, L. (2023). Exploring participant attrition in a longitudinal follow-up of older adults: the Global Longitudinal Study of Osteoporosis in Women (GLOW) Hamilton cohort. *BMJ Open* 13:e066594. doi: 10.1136/bmjopen-2022-066594
- Petersen, G. L., Jorgensen, T. S. H., Mathisen, J., Osler, M., Mortensen, E. L., Molbo, D., et al. (2024). Inverse probability weighting for self-selection bias correction in the investigation of social inequality in mortality. *Int. J. Epidemiol.* 53:dyae097. doi: 10.1093/ije/dyae097
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Raciborski, F., Klak, A., Kwiatkowska, B., Batko, B., Sochocka-Bykowska, M., Zon-Giebel, A., et al. (2017). Diagnostic delays in rheumatic diseases with associated arthritis. *Reumatologia* 55:29056771, 169–176. doi: 10.5114/reum.2017.69777
- Sjoberg, D. D., Curry, M., Lavery, J. A., and Larmarange, J. (2021). Reproducible summary tables with the gtsurvey package. *R J.* 31, 570–580. doi: 10.32614/RJ-2021-053
- Sperrin, M. R., Richard, D., Collins, G. S., and Martin, G. P. (2022). Targeted validation: validating clinical prediction models in their intended population and setting. *Diagn. Progn. Res.* 6:24. doi: 10.1186/s41512-022-00136-8
- Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., et al. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 338:b2393. doi: 10.1136/bmj.b2393

- van Buuren, S. (2018). *Flexible Imputation of Missing Data*. 2nd Edn Boca Raton., MA: CRC Press.
- Wallace, W., Chan, C., Chidambaram, S., Hanna, L., Iqbal, F. M., Acharya, A., et al. (2022). The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. *npj Digit Med.* 5:118. doi: 10.1038/s41746-022-00667-w
- Weir, C. B., and Jan, A. (2024). "BMI classification percentile and cut off points," in *StatPearls*, (Treasure Island (FL): StatPearls Publishing).
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Winn, A. N., Somai, M., Fergestrom, N., and Crotty, B. H. (2019). Association of use of online symptom checkers with patients' plans for seeking care. *JAMA Netw. Open* 2:e1918561. doi: 10.1001/jamanetworkopen.2019.18561
- Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., et al. (2019). PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann. Intern. Med.* 170, 51–58. doi: 10.7326/m18-1376
- Wu, M. J., Zhao, K. L. Y., and Fils-Aime, F. (2022). Response rates of online surveys in published research: a meta-analysis. *Comput. Hum. Behav. Rep.* 7:100206. doi: 10.1016/j.chbr.2022.100206
- Zhang, X., Holt, J. B., Lu, H., Wheaton, A. G., Ford, E. S., Greenlund, K. J., et al. (2014). Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *Am. J. Epidemiol.* 179, 1025–1033. doi: 10.1093/aje/kwu018