

VIEWPOINT

Percentage of progressors in imaging:
can we ignore regressors?Alexandre Sepriano,^{1,2} Sofia Ramiro,^{1,3} Robert Landewé,^{3,4} Maxime Dougados,⁵
Desirée van der Heijde¹

To cite: Sepriano A, Ramiro S, Landewé R, *et al*. Percentage of progressors in imaging: can we ignore regressors?. *RMD Open* 2019;5:e000848. doi:10.1136/rmdopen-2018-000848

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/rmdopen-2018-000848>).

Received 30 October 2018

Revised 30 January 2019

Accepted 31 January 2019

ABSTRACT

Stopping or preventing structural progression is a goal common to all inflammatory rheumatic diseases. Imaging may capture structural progression across diseases, but is susceptible to measurement error. Progression can be analysed as a continuous change score over time (eg, mean change of the van der Heijde-modified Sharp score) or as a binary change score (eg, percentage of progressors according to the modified New York criteria). Here, we argue that the former takes measurement error into account while the latter ignores it, which may lead to spurious conclusions. We will argue that assumptions underlying commonly used binary definitions of progression are false and we propose a method that incorporates (inevitable) measurement error.

Inflammatory rheumatic musculoskeletal diseases (RMDs), such as rheumatoid arthritis (RA) and spondyloarthritis (SpA), typically cause irreversible joint damage over time, particularly if left untreated. Recent landmark therapeutic advancements suggest modifying the destructive course of a disease is possible, but still much needs to be done in this regard.^{1,2} In order to capture treatment effects in joint damage progression, valid outcome measures are warranted, as prescribed by regulatory agencies worldwide.³⁻⁶

Conventional radiography is the standard modality for capturing and quantifying progression of structural damage in RMDs. Although we focus on conventional radiography as an example, the issues we address here apply similarly to all imaging modalities assessing structural damage. Equally important as the imaging modality itself is the analytical method used to quantify progression. For example, radiographic progression can be analysed as an averaged continuous change score (eg, mean change of the van der Heijde-modified Sharp score [SvdH] over time; or the modified Stoke Ankylosing Spondylitis Spine Score [mSASSS] over time) or as a binary change score (eg, percentage of 'progressors' according to the modified

New York criteria [mNY]). Another way of presenting a binary change score is dichotomising a continuous change score (eg, SvdH ≥ 5 vs < 5 ; or mSASSS ≥ 2 vs < 2). The quantification of radiographic progression, like outcome assessments in general and other imaging methods more specifically, is susceptible to measurement error. Here, we will demonstrate that researchers using continuous change scores will implicitly take measurement error into account, while researchers using binary change scores will frequently omit measurement error.

We make a plea that measurement error (or noise) should not be ignored when interpreting imaging studies. The 'signal-to-noise' ratio analogy has been recently proposed to better explain the fallacies of ignoring measurement error.⁷ Here, this analogy will be used to argue the false assumptions underlying commonly used binary definitions of progression. The 'signal-to-noise' concept incorporates two types of information: (1) 'true change' ('signal') and (2) error change ('noise'). The larger the measurement error, the harder to capture the 'signal' and in some cases, disentangling the 'signal' from the 'noise' can be particularly challenging. Sources of 'noise' in reading radiographs are plenty and widely recognised (eg, technical, intra-reader and inter-reader variability). To improve the 'signal-to-noise' ratio (the higher the better), investigators have been implementing strategies to reduce the denominator (ie, 'noise') by, for instance, combining judgements from ≥ 2 trained central readers. Nevertheless, these (methodological) strategies cannot fully eliminate the undesired 'noise'. Thus, here we discuss how appropriate analytical choices can further contribute to handle 'noise' in imaging assessment, ultimately contributing to its reduction.

We have used data from a recently published study from the DEvenir des Spondylarthropathies Indifférenciées Récentes (DESIR)



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

²NOVA Medical School, Universidade Nova de Lisboa, Lisbon, Portugal

³Department of Rheumatology, Zuyderland Medical Center, Heerlen, The Netherlands

⁴Amsterdam Rheumatology Center, AMC, Amsterdam, The Netherlands

⁵Hopital Cochin, Rheumatology, Université Paris Descartes, Paris, France

Correspondence to

Dr Alexandre Sepriano;
alexsepriano@gmail.com

cohort⁸ to better illustrate the concept of ‘signal-to-noise’ ratio with a particularly challenging case. In our example, damage occurring in the sacroiliac joints (SIJs) over 5 years was evaluated in patients with axial SpA (axSpA), according to the mNY scoring system.⁹ This scoring system has clearly been shown to be unreliable (much ‘noise’), especially if scores from only one (untrained) reader are used.^{10–12} We reduced the ‘noise’ by having baseline and 5-year films per patient scored by three trained central readers obtained independently, and used blinded chronological order to ensure unbiased measurement error in two directions (ie, the readers did not know which is the baseline and which is the 5-year film when scoring the pair). Each reader reported a binary score (mNY-positive vs mNY-negative) and a (semi)continuous grade (range, 0–8; both SIJ together) per time point. The final mNY binary status score was defined by the agreement of at least two of the three readers, and the continuous grade by the average of the three independent scores. The binary change scores can take three possible values (–1, 0, +1). For instance, if a patient is mNY-positive at baseline and negative at 5 years, the binary change score is –1 (negative change or ‘improvement’). Similarly, the continuous change score can also be positive and negative (range, –8 to +8), where a negative value means the mNY grade at 5 years is smaller than the grade at baseline. The resulting change scores are shown here in a way that makes measurement error better visible: (1) for the binary change score, we show the cross-tabulation between baseline and 5-year combined scores (table 1), and report positive change (ie, worsenings; +1) and negative change (ie, ‘improvements’; –1); and (2) for the continuous change score, we report a cumulative probability plot (figure 1) that (by default) also shows positive and negative change and, additionally, we overlay the binary changes in the plot to facilitate comparison. These data are used here as the ‘common ground’ from which we explore the assumptions of commonly used binary definitions of progression, and finally to propose an assumption-free approach. This is all under the assumption that structural damage is irreversible (which might not necessarily apply in all settings) and therefore improvements should be judged as measurement error.

Table 1 Change in the mNY status in patients with axSpA after 5 years in the DESIR cohort⁸

Baseline	5 years		
	mNY positive	mNY negative	Total
mNY positive	59	3	62
mNY negative	24	330	354
Total	83	333	416

axSpA, axial spondyloarthritis; mNY, radiographic sacroiliitis according to the modified New York criteria (agreement between ≥ 2 out of 3 trained central readers blinded to time order).

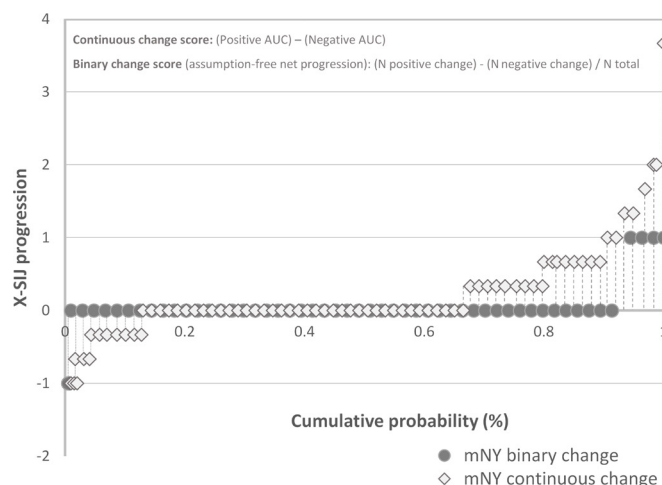


Figure 1 Cumulative probability plot. Structural progression in radiographs of the sacroiliac joints (X-SIJ) according to the modified New York criteria (mNY), measured as binary change (possible values: +1, 0, –1) and continuous grade change (range of possible values: –8; 8). Each data point represents either binary (black circles) or continuous (open diamonds) progression from one unique patient (selection from the total sample to increase readability but covering the full range of observed values). Positive AUC: dashed lines; negative AUC: dotted lines. AUC, area under the curve; N, number of patients.

CRUDE PROGRESSION

At baseline, 62 (15%) of the 416 patients were classified as mNY-positive. Of the 354 mNY-negative patients at baseline, 24 changed into mNY-positive after 5 years (positive change or worsening; +1). Most studies would have only reported these 24 cases (6.8% [24/354]) as those who had progressed from mNY-negative to mNY-positive.^{13–16} But this rate is spuriously high for two reasons: First, it implies that the baseline reading is true and free of measurement error (bias); second, it assumes that a change in the unexpected opposite direction (negative change or ‘improvement’; –1) can be ignored. Since radiographic readings are not free of measurement error and readers are not aware of which film pertains to baseline and follow-up, such an approach does not provide a valid representation of the truth. Also when analysing the data, one must consider the different possible scenarios, in this case meaning that ‘improvement’ or negative change, though less expected or warranted, can also happen, particularly due to measurement error. This method to measure progression does not accommodate this reality.

CONDITIONAL NET PROGRESSION

Recently, researchers from the DESIR and the German Spondyloarthritis Inception Cohort reported progression of radiographic sacroiliitis at 2 years.^{17 18} They acknowledged that a robust estimation of progression must not ignore the measured negative changes. Table 1 shows how this principle worked out: Positive changes (‘worsening’ in 24 of the 354 formerly mNY-negative patients [6.8%]; ‘+1 change’) and negative changes (‘improvement’ in

3 of the 62 formerly mNY-positive patients [4.8%]; ‘-1 change’) were seen, and ‘net progression’ was obtained by calculating the difference between both rates (2%). While this approach differs from the ‘crude method’ by acknowledging the relevance of negative changes, the ‘net progression’ rate of 2% is still conditional on the baseline classification status assumed to be free of bias. In other words, it implicitly assumes that ‘worsening’ can only happen in patients who are mNY-negative at baseline and ‘improvement’ only in mNY-positive patients. Since readers are not aware which film is the baseline film (scores had been obtained in pairs with full blinding of time order), this assumption does not hold.

ASSUMPTION-FREE NET PROGRESSION

We therefore propose an assumption-free method to analyse structural damage progression.⁸ In principle, both ‘positive changes’ (+1 change) and ‘negative changes’ (-1 change) are ‘allowed’ and scores of individual patients are not interpreted as ‘true progression’ or ‘noise’. Under the premise of reading with concealed (blinded) time order, measurement error (‘noise’) presumably occurs symmetrically. This means that it will affect scores with similar likelihood in both directions since readers are not aware of which image pertains to baseline and which to follow-up, as has been worked out by us previously for progression in RA.¹⁹ So, with the ‘assumption-free’ method, the overall improvement contains (in theory) both ‘true improvement’ (ie, repair) as well as measurement error. Similarly, worsening also includes ‘true worsening’ (ie, progression) and measurement error. However, in a setting of irreversible damage, it is not unreasonable to expect that measurement error (rather than repair) largely dominates improvements. Still the direction and magnitude of residual bias (driven by bidirectional measurement error) is difficult to know with certainty for binary outcomes. Notwithstanding with the proposed method measurement error at least is incorporated and not ignored as done thus far.

With the ‘assumption-free’ method, if ‘true progression’ is present over-and-above measurement error, it will become obvious as a positive change when all zero changes, positive changes and negative changes occurring in the *entire* population are summed together. The area under the curve (AUC) of the probability plots (positive area minus negative area) provides the mean continuous change score taking measurement error into account since it incorporates, by default, both positive (>0, ie, corresponding to ‘+1 change’) and negative (<0, ie, corresponding to ‘-1 change’) changes (figure 1). In our example, the overall mean change-score (+0.20 [SD 0.55]) can be obtained by the subtraction of the mean status score at baseline (1.40 [SD 1.68]) from the mean status score at 5 years (1.60 [SD 1.83]). Another way of getting the average continuous change score is by summing all positive change-scores (+106.67 N=136 within the positive AUC), all negative change-scores (-24.67; N=53

within the negative AUC) and all no-changes (0; N=227) and divide the result by the total number of patients [(10 6.67+(-24.67)+0)/416=+0.20]. Thus, on average, the continuous change score is positive (+0.20) since positive change scores outweigh the negative change scores but, importantly, both are included in the calculation. The binary ‘assumption-free’ net progression is analytically similar, also capturing measurement error appropriately. However, measurement error is neglected by the first two definitions of binary change. If positive binary changes are scored +1, negative changes are scored -1, and no changes are scored zero, the total change is the sum of all +1 scores, -1 scores and zeros scores, divided by the total number of observations, and expressed as a percentage [(24+(-3)+0)/416=5%]. Similar to the average continuous change score above (+0.20), an overall positive percentage implies that, at the group level, there is more progression than measurement error. By doing so, we get an ‘assumption-free’ net progression of +5% and not of +2% (as the conditional net progression).

Of note, the estimated progression is an averaged estimate which aims to approximate ‘true progression’ at the group level (ie, beyond measurement error) but does not translate to individual patients. So, it becomes impossible to declare a patient as a ‘progressor’, as is often done in the context of clinical trials. Similarly, we estimate 5% progression from mNY-negative to mNY-positive after 5 years in the population of DESIR patients, and not 21 progressors out of 416.

PROPOSED METHOD FOR FUTURE RESEARCH

In summary, three methods to approximate binary progression to ‘true progression’ that are in use have been discussed here: (1) ‘crude progression’; (2) ‘conditional net progression’; (3) ‘assumption-free net progression’. This ‘assumption-free net progression’ yields the least biased estimates since it gives most credit to measurement error (ie, always includes error without prior assumptions on the imaging modality ability to reliably capture change or on the baseline status score). Obviously, decreasing bias carries many benefits such as the better detection of treatment effects in randomised trials. Thus, we propose that this method will be applied in future studies with binary imaging outcomes. Importantly, this method applies to both continuous outcome measures that are dichotomised (eg, SvdH ≥ 5 vs <5; or mSASSS ≥ 2 vs <2) as well as to dichotomous measures by nature (eg, mNY-positive vs mNY-negative),^{9,20} but should be used with caution since it implies that outcomes are irreversible (mainly structural damage), and are evaluated over not too long periods, as ‘true repair’ cannot be excluded with longer follow-up. A better understanding of what structural repair means (and importantly how to define it) is still a major unmet need in the field of rheumatology. Further studies are necessary to better understand ‘negative changes’ in settings other than irreversible damage and how ‘true improvements’ (ie,

repair) possibly contribute to the overall net progression. However, since the proposed ‘assumption-free’ method, different to what has been done so far, implies full disclosure of the bidirectional change (eg, as a 2:2 table used in this viewpoint), together with the overall figure of ‘net progression’, it can facilitate research pursuing a consensual definition of ‘repair’ by acknowledging and, importantly, making ‘negative change’ more visible. This includes subtle distinctions between, for instance, spontaneous repair and repair driven by interventions which might reflect different pathophysiological pathways. Understanding these differences will allow a better interpretation of the treatment effects of drugs targeting specific pathways and how the ‘assumption free’ method captures these effects.

While we have used the example of radiographs in axSpA, the application of assumption-free net progression extends to all examples in rheumatology where imaging scores on structural damage are obtained under blinded conditions, and likely goes beyond. The example of axSpA should here be merely seen as an example of a methodological issue that we would welcome researchers to incorporate in their analysis of radiographic progression, independently of the disease being investigated. Too often we think that measurement error is not a big issue, while it is really there but often only not quantified.

Contributors All authors gave substantial contributions to the conception of the paper as well as drafting the paper and revising it critically for important intellectual content.

Funding DESIR is financially supported by an unrestricted grant from Pfizer. AS is supported by a doctoral grant from ‘Fundação para a Ciência e Tecnologia’ (SFRH/BD/108246/2015).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Smolen JS, Landewé R, Bijlsma J, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 2017;76:960–77.
- van der Heijde D, Ramiro S, Landewé R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis* 2017;76:978–91. Jun;.
- European Medicines Agency. Concept paper on clinical investigation of medicinal products for the treatment of axial spondyloarthritis. Available: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/04/WC500185187.pdf
- European Medicines Agency. Guideline on clinical investigation of medicinal products for the treatment of rheumatoid arthritis. Available: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2018/01/WC500241042.pdf
- Food and Drug Administration. Clinical trial imaging endpoint process standards guidance for industry. Available: <https://www.fda.gov/downloads/drugs/guidances/ucm268555.pdf>
- Boers M, Brooks P, Strand CV, et al. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198–9.
- Landewé RBM, van der Heijde D. “Big Data” in Rheumatology: Intelligent Data Modeling Improves the Quality of Imaging Data. *Rheum Dis Clin North Am* 2018;44:307–15.
- Dougados M, Sepriano A, Molto A, et al. Sacroiliac radiographic progression in recent onset axial spondyloarthritis: the 5-year data of the DESIR cohort. *Ann Rheum Dis* 2017;76:1823–8.
- van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984;27:361–8.
- Sepriano A, Rudwaleit M, Sieper J, et al. Five-year follow-up of radiographic sacroiliitis: progression as well as improvement? *Ann Rheum Dis* 2016;75:1262–3.
- van den Berg R, Lenczner G, Feydy A, et al. Agreement between clinical practice and trained central reading in reading of sacroiliac joints on plain pelvic radiographs. Results from the DESIR cohort. *Arthritis Rheumatol* 2014;66:2403–11.
- van Tubergen A, et al. Radiographic assessment of sacroiliitis by radiologists and rheumatologists: does training improve quality? *Ann Rheum Dis* 2003;62:519–25.
- Schattenkirchner M, Krüger K. Natural course and prognosis of HLA-B27-positive oligoarthritis. *Clin Rheumatol* 1987;6(Suppl 2):83–6.
- Mau W, Zeidler H, Mau R, et al. Outcome of possible ankylosing spondylitis in a 10 years’ follow-up study. *Clin Rheumatol* 1987;6(Suppl 2):60–6.
- Sampaio-Barros PD, Conde RA, Donadi EA, et al. Undifferentiated spondyloarthropathies in Brazilians: importance of HLA-B27 and the B7-CREG alleles in characterization and disease progression. *J Rheumatol* 2003;30:2632–7.
- Sampaio-Barros PD, Bortoluzzo AB, Conde RA, et al. Undifferentiated spondyloarthritis: a longterm followup. *J Rheumatol* 2010;37:1195–9.
- Poddubnyy D, Rudwaleit M, Haibel H, et al. Rates and predictors of radiographic sacroiliitis progression over 2 years in patients with axial spondyloarthritis. *Ann Rheum Dis* 2011;70:1369–74.
- Dougados M, Demattei C, van den Berg R, et al. Rate and predisposing factors for sacroiliac joint radiographic progression after a two-year follow-up period in recent-onset spondyloarthritis. *Arthritis Rheumatol* 2016;68:1904–13.
- Landewé R, van der Heijde D. Radiographic progression depicted by probability plots: presenting data with optimal use of individual values. *Arthritis Rheum* 2004;50:699–706.
- van der Heijde D. How to read radiographs according to the Sharp/van Der Heijde method. *J Rheumatol* 2000;27:261–3.