

R-index: a standardized representativeness metric for benchmarking diversity, equity, and inclusion in biopharmaceutical clinical trial development



Spencer L. James,^{a,b,*} Max Bourgognon,^{c,d} Patricia Pinto Vieira,^{c,d} Bruno Jolain,^{c,d} Sarah Bentouati,^a Emma Kipps,^e Assaf P. Oron,^f Catherine W. Gillespie,^f Ruma Bhagat,^a Altovise Ewing,^a Shalini Hede,^a Keith Dawson,^a and Nicole Richie^a



^aGenentech Inc., 1 DNA Way, South San Francisco, CA, 94080, USA

^bDocere Research, 1037 NE 65th St, Seattle, WA, 98115, USA

^cRoche UK, Hexagon Place, Shire Park, Falcon Way, Welwyn Garden City, AL7 1TW, UK

^dF. Hoffmann-La Roche Ltd, Unterstrass 124, 4070, Basel, Switzerland

^eNational Health Service England — London, Wellington House, 133-135 Waterloo Road, London, SE1 8UG, UK

^fInstitute for Health Metrics and Evaluation at the University of Washington, 3980 15th Avenue NE, Seattle, WA, 98195, USA

Summary

Background Diversity, equity, and inclusion pertaining to race, ethnicity, and related concepts have historically been underrepresented in clinical trials for pharmaceutical drug development, although this is an increasing topic for regulators, payers, and patient advocacy groups. We aimed to develop a summary statistical measure to assess such representativeness.

Methods A statistical measure using population demographic parameters derived from performance metrics through verbal autopsy research was proposed for using population frameworks in the UK. The summary measure, R-index, was demonstrated using simulation data with population frameworks from the UK (116 Roche UK clinical trials 2013–2022) and then using published clinical trial results (NCT02366143 [March 1, 2015–September 15, 2017], NCT04368728 [July 27, 2020–October 9, 2020], and NCT04470427 [July 27, 2020–November 25, 2020]). R-index was further proposed for use with benchmarking performance in representative trial development for internal processes, external benchmarking, and performance tracking in clinical trial development.

Findings R-index was derived from a standardized statistical measure called the L1 norm, or Manhattan distance, and then normalized to the maximum theoretical error observed in some populations using population framework or ontology for reporting concepts such as race, ethnicity, and other dimensions of diversity used to characterize patient cohorts. R-index demonstrated desirable qualities in demonstration simulations, including a range of 0–1, ease of calculation and use, and interpretability and flexibility, as data standards in the space of inclusive research continue to develop.

Interpretation R-index is an interpretable, accessible summary statistic that may be useful for tracking and benchmarking representativeness in inclusive research and related domains. R-index is adaptable to different population frameworks and ontologies across different settings and considerations in terms of underlying population variables.

Funding F. Hoffmann-La Roche Ltd/Genentech, Inc.

Copyright © 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Representativeness; Clinical trials; Advancing inclusive research

Introduction

For ethical and scientific reasons, clinical research must be representative of the population that experiences a disease. Representativeness in this context means that a clinical research study should not over- or under-enroll

a subpopulation relative to the population with disease unless there is scientific or statistical basis. Achieving representation in clinical trials can be complex given the multiple factors that affect disease risk, severity, and detection and the likelihood that a patient with the

*Corresponding author. PO Box 2099, El Granada, CA, 94018-8055, USA.
E-mail address: spencj@uw.edu (S.L. James).

Research in context

Evidence before this study

Representativeness is a critical concept in clinical trials and drug development, yet a metric for measuring representativeness has not been defined, standardized, or adopted for research practices. Existing methods and metrics are not commonly used and do not necessarily address the unique challenge of incorporating categorical variables such as race and ethnicity, for which distributions need to approximate some reference population.

Added value of this study

This study proposes a novel metric, termed R-index, that summarizes the statistical differences in two Dirichlet distributions, each of which sums to 100%. R-index has desirable properties: it uses a scale between 0 and 1; it is

calculated using enrollment data from a clinical trial and population baseline data considered to be the reference for representative enrollment in that setting; and it is a univariate measure that serves as a summary statistic between and across trials, as opposed to an individual error metric that may not be comparable across different frameworks (eg, between US and UK census data collection standards).

Implications of all the available evidence

We propose the R-index as a standardized metric for summarizing and reporting representativeness of race, ethnicity, and other diversity-related attributes of clinical trial participants that can be used and translated across biopharmaceutical research and development, regardless of therapeutic area and trial phase.

disease will enroll in a clinical study. Disease risk and severity may be influenced by genetic, environmental, behavioral, and socioeconomic factors that correlate with barriers to diagnosis and care. This suggests that the propensity for a patient with a disease to enter a clinical trial may be further affected by factors beyond the risk of disease, including the probability of diagnosis; access to and trust of the healthcare system; literacy, including health literacy; language barriers; and provider and societal factors, including racism, sexism, and other forms of systemic bias.^{1–4}

There is increasing consensus among scientists, regulators, patients, and advocacy groups that research must be representative of patients with diverse demographics, including social descriptors of race and ethnicity and biological characteristics of genetic ancestry, yet standardized metrics for assessing and reporting representativeness are not clearly defined. While there are examples of statistical measures that address this concept, such as transportability between trials and populations, recent studies such as those for Covid-19 vaccines did not report extensively on representativeness of participants beyond descriptive statistics, which often lacked data on population-specific disease burden.⁵ However, representativeness is a critical concept to estimate the extent to which an experiment's results can be expected in the population that stands to benefit, indicating a need for representativeness in clinical research to be benchmarked, compared, and tracked over time and across different settings in terms of disease, therapeutics, and patient population. This aligns with how representativeness is increasingly being considered a key component in assessing the robustness of a clinical trial alongside conventional statistics such as power and sample size, which are frequently evaluated by regulatory agencies and discussed in clinical practice. While representativeness is a moral and scientific imperative in itself, as a measurable

entity it could also help organizations prioritize and measure progress in diversity, equity, and inclusion in research settings.³

One of the challenges in measuring representativeness is the increasing detail with which human diversity itself is characterized in terms of genomics, race, ethnicity, ancestry, sex, gender, and other variables and combinations thereof. Ontologies used for recording demographics and related information over the past century have limited granularity compared with modern standards. It is likely that the frameworks used for reporting the results of the 2020 US Census will change by the next census to add further detail, and these shifting frameworks pose challenges for how scientific research can measure representativeness as it pertains to key variables currently recognized as well as future variables. Cross-location comparisons of representativeness can be limited given differences in ontologies and frameworks for characterizing populations even when they are addressing similar concepts. For instance, the 2020 US Census documentation from the Office of Management and Budget specified the following for reporting race:⁶ *White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or other Pacific Islander*. By comparison, the 2021 UK Census documentation specified the following for categorizing ethnicity:⁷ *Asian or Asian British, which included Indian, Pakistani, Bangladeshi, Chinese, and any other Asian background; Black, Black British, Caribbean, or African, which included Caribbean, African, and any other Black, Black British, or Caribbean background; mixed or multiple ethnic groups, which included White and Black Caribbean, White and Black African, White and Asian, and any other mixed or multiple ethnic background; White, which included English, Welsh, Scottish, Northern Irish or British (in Wales, Welsh is the first option in the White category), Irish, Gypsy or Irish Traveller, Roma, and any other White background; and*

other ethnic group, which included Arab and any other ethnic group.

These frameworks represent fundamentally different ontologies for recording race and ethnicity. Similar comparisons could be drawn longitudinally over time as ontologies evolve. Furthermore, it would be expected that governmental agencies, health professionals, researchers, and regulatory officials across locations representing different interests, geographies, and populations would have additional frameworks relevant for data collection and analytical purposes. From this example comparing the US and the UK, it is evident that a specific framework developed for US biopharmaceutical research may have limited application in the UK and vice versa. Sample size or study design characteristics may be used to qualitatively appraise the validity of one result over another; however, representativeness with respect to some defined population distribution as a quantitative entity is not rigorously defined in a way that permits comparability unless two studies are nearly identical.

This paper proposes a representativeness index (R-index) for measuring representativeness in clinical trials and related interventions with a derivation that allows for its tracking and benchmarking across space, time, and setting without requiring universal consensus on ontologies. The concept can be applied retrospectively, in real time, and for simulation and counterfactual forecasting. Calculating the R-index is possible using descriptive statistics in epidemiological and demographic data without complex software or statistical models. R-index requires estimates of disease burden parameters at the demographic level and clinical trial enrollment data with an equivalent or mapped demographic ontology. We demonstrate use of the metric in simulated and real data from UK clinical trials and selected examples from published literature.^{8–10} This metric is proposed as a standard reporting measure for use in clinical trial development to advance diversity, equity, and inclusion as scientific imperatives in biopharmaceutical research.

Methods

UK clinical trial data and catchment area

Data from completed Roche UK clinical trials between 2013 and 2022 were analyzed ([Supplemental Table S1](#)). A total of 5288 patients who participated in these clinical trials (phase I–IV, across all therapy areas) and had ethnicity recorded were included using a clinical trial repository. These patients represented approximately one-third of patients recruited over the 10 years. All patients provided written informed consent. Missing data were mainly attributed to the lack of ethnicity coding (manual input) and the partial coverage of studies in our repository. Variables available from the National Health Service (NHS) Trust centers were trial

start and end dates, disease and therapeutic area, and enrollment by the following ethnic groups (the term *ethnicity* is used to align with UK terminology, recognizing that in US frameworks, these may be termed *race*): Asian; Black; White, American Indian or Alaska Native; White, Black; White, Asian; White, other; other, White; Black, White; Asian, White; unknown, other; American Indian or Alaska Native; Native Hawaiian and other Pacific Islander; and White.

The UK clinical trial data included statistics on enrolled patient populations by trial and NHS Trust site. For each trial, the number of patients enrolled by ethnicity at each site were identified. For each site, estimates of the catchment area population by ethnicity were provided directly by each NHS Trust where a trial took place.¹¹ Categorization of ethnicity values was different in the Roche UK trial data, which had more granular categories than the NHS catchment population data, which relied on broader UK Census categories from 2000. A map was developed to align the more granular trial ethnicity categories with the broader NHS and census ethnicity categories and was reviewed by internal experts in the UK to align with the categories and subcategories defined with 2011 UK Census standards.¹²

The clinical trial and catchment data allowed measurement of the proportions of patients with a disease who belonged to one of five ethnicity groups (Black, White, Asian, mixed, other); absolute error between an NHS Trust's catchment area and its historical trial enrollment population on a site trial-specific basis was analyzed. The R-index was adapted from an analogous problem with different parameters proposed by Murray et al.¹³ in comparing cause-specific mortality fraction values in verbal autopsy classification of causes of death in population health research. Representativeness was summarized on a scale of 0–1, with 0 having the highest possible error given the population distribution and 1 having zero error and complete representativeness across a specified ontology (the UK ethnicity reporting framework).

US clinical trial data and catchment area

The R-index was calculated for the US IMpower150 trial (NCT02366143) in metastatic non-small cell lung cancer (NSCLC) in order to demonstrate its use in a different geographical and disease area setting.¹⁰ IMpower150 was chosen as it was a Roche study that had included demographic descriptions in its publication. The published results were used for proof-of-concept purposes to demonstrate how R-index can be calculated from summary results. Baseline NSCLC prevalence data from the Surveillance, Epidemiology, and End Results (SEER) Program were compared with the published demographic data from IMpower150, and absolute error was calculated in the proportion by race and ethnicity group. In IMpower150, Hispanic and non-Hispanic

categories were not delineated or reported separately in descriptive statistics. The R-index was calculated using race as reported. Due to categorizations used in IMpower150 and in SEER, post hoc aggregates were developed to convert the tabulations to comparable groups.

Additional clinical trial data and catchment area

Clinical trials that were conducted at different global sites with available comparator data for demographic groups were identified, with a decision to assess R-index for two studies focused on Covid-19 vaccines that were highly visible during the height of the Covid-19 pandemic. The R-index computation was applied to the article “Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine,” which reported efficacy of the first approved mRNA vaccine for Covid-19 in a study conducted in the US, Argentina, Brazil, and South Africa.⁸ Distributions of race and ethnicity were extracted from the article, and comparable proportions of the US population using 2020 US Census data were used to estimate the absolute error between proportions of each race and ethnicity in the trial and US populations. The article “Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine,” which reported on a study conducted in a US population, was also analyzed.⁹ Demographic statistics were extracted from the article and the same population proportions were used to estimate the R-index.

Formulation

The problem in measuring representativeness using multinomial, categorical race and ethnicity variables is quantifying the dissimilarity between two Dirichlet distributions, A (population) and B (trial), in statistical terms. Dirichlet distributions represent categorical probability values that sum to 100%. The challenge lies in developing a metric that captures the magnitude of errors while normalizing the measure to the scale of one distribution, enabling standardized comparisons between the two distributions. This proposed formula was adapted from Murray et al. using different parameters from statistical performance metric research used in population health studies through verbal autopsy, a method of collecting cause-of-death information in areas that lack vital registration systems.¹³ The standardization of metrics has helped standardize and compare verbal autopsy methods and instruments to improve performance.

The R-index is computed as 1 minus the quotient of the sum of absolute errors and the maximum possible error between two Dirichlet distributions. To derive the R-index, we first considered the L1 norm, or Manhattan distance, which measures the absolute differences between the corresponding components of two probability vectors, which must have values between 0 and 1 inclusive. The L1 norm provides a basis for capturing the

dissimilarity between the distributions. We modified this distance metric by normalizing it to the maximum possible error based on the population (“ground truth”) distribution, resulting in a range-bound metric.

The R-index is calculated by the following formula:

$$R\ index_{population}^{trial} = 1 - \frac{\sum_{j=1}^J |Proportion_j^{population} - Proportion_j^{trial}|}{2 \times (1 - \min(Proportion^{population}))}$$

for J demographic groups, in which proportion is defined as the number of people with the disease in demographic group j divided by all people with the disease in the entire population.

The maximum possible error is obtained by identifying the category in the population distribution with the lowest proportion and setting the proportion of that category to 100% in the other trial enrollment. This yields the maximum possible discrepancy between the distributions. R-index retains the desirable properties of L1 norm, including non-negativity and symmetry. This ensures that the R-index captures positive and negative differences between distributions while treating differences in either direction equally, recognizing that in real-world clinical trials, there may be a need for contextual, nuanced considerations of disease area, therapeutic area, population, and healthcare sector. The resulting metric is bounded between 1 and 0, with 1 indicating identical distributions and 0 representing maximal dissimilarity.

Demonstration

An example was developed with simulated data to demonstrate how R-index is computed and interpreted using shifts in the simulated trial populations for five ethnicity groups in the UK trials (2011 UK Census data: overall; [Table 1](#)). These ethnicity groups were adopted from the 2011 UK Census data and underwent revision for the 2021 UK Census. The R-index value was computed as the following: $1 - (0.05 + 0.05 + 0.05 + 0.05 + 0.10) / (2 \times (1 - 0.10)) = 0.833$. The same population but from a trial with zero error was considered (2011 UK Census data: zero error; [Table 1](#)). If there was zero error in the trial with the proportion in the population equaling the proportion in the trial for every group, then the numerator equaled 0 and the R-index equaled 1.00. If there was error in this dataset (eg, if only White patients were enrolled with the same baseline population distribution), then the R-index would yield a value of $1 - (0.20 + 0.70 + 0.10 + 0.20 + 0.10) / (2 \times (1 - 0.10)) = 0.278$, worse than the initial scenario (2011 UK Census data: more error; [Table 1](#)).

Role of the funding source

The funder of the study had a role in study design, data collection, data analysis, data interpretation, and writing of the report.

2011 UK Census data: overall	Proportion (population)	Proportion (trial)	Absolute error
Black	0.20	0.25	0.05
White	0.30	0.35	0.05
Asian	0.10	0.05	0.05
Mixed	0.20	0.15	0.05
Other	0.10	0.20	0.10
2011 UK Census data: zero error			
Black	0.20	0.20	0.00
White	0.30	0.30	0.00
Asian	0.10	0.10	0.00
Mixed	0.20	0.20	0.00
Other	0.10	0.10	0.00
2011 UK Census data: more error			
Black	0.20	0.00	0.20
White	0.30	1.00	0.70
Asian	0.10	0.00	0.10
Mixed	0.20	0.00	0.20
Other	0.10	0.00	0.10

Table 1: Population and trial enrollment statistics for five ethnicity groups in the UK trials.

Results

UK clinical trial data

A total of 116 UK clinical trials with 1443 patients were identified after removing trials with ≤ 2 sites and missing ethnicity data (Supplemental Table S1). Since the parameters required only estimates for expected and enrolled populations, R-index was calculated for each UK trial-study site combination (Fig. 1), across disease areas (Fig. 2), for different NHS Trust sites (Fig. 3), and across time (Fig. 4). Baseline data on the catchment area population, which were not disease specific, were provided by NHS Trust estimates.

US clinical trial data: IMpower150

Table 2 shows the patient counts and proportions from IMpower150 and SEER NSCLC data by race; the R-index was 0.7797. If another 150 patients were enrolled, for a total enrollment of 1000, and the investigators aimed to enroll these 150 patients to maximize representativeness using the R-index, optimizing the R-index using all five groups would be an option. However, the table shows that the Asian and Hawaiian/Pacific Islander groups and other, multiple, and unknown group were overrepresented relative to baseline. The American Indian or Alaska Native proportion was approximately equal to that at baseline. For the two underrepresented groups, White (A) and Black or African American (B), in which $A + B$ should equal 150 and the R-index was similarly calculated with a total N of 1000, enrolling 93 additional Black or African American patients and 57 additional White patients would increase the R-index to 0.821.

Additional clinical trial data

The proportions of race and ethnicity in the BNT162b2 mRNA Covid-19 vaccine trial, estimates from the 2020 US Census population, and computed error are shown in Table 3. The R-index was computed as follows:

$$1 - (0.213 + 0.031 + 0.017 + 0.006 + 0.000 + 0.163 + 0.005) / (2 \times 1 - (0.001)) = 0.782$$

Descriptive statistics from the mRNA-1273 SARS-CoV-2 vaccine trial and estimates from the 2020 US Census population are shown in Table 3. The R-index was 0.816.

Discussion

This study proposes a metric for assessing trial representativeness in the biopharmaceutical research setting. R-index summarizes the representativeness of a trial's enrollment population compared with the population of interest using a demographic framework, which is likely to continue evolving. While this study demonstrated the use of R-index for a limited number of racial and ethnic categories, it can be applied broadly to other variables such as gender, which is a more complex categorical variable than the traditional binary of male or female sex. The purpose of the R-index is to allow drug developers, regulators, and others to assess the representativeness of a trial in dimensions of diversity that are relevant to health equity, genomics, and ancestry or otherwise. This framework and standardization may be useful as discourse on advancing inclusive research evolves.

Our demonstration of the R-index included a theoretical application using a simplified model dataset, calculation of the R-index for historical UK clinical trial data and select US clinical trials, and example calculations for two US mRNA Covid-19 vaccine studies. In the absence of post hoc reporting and calculation across historical trials from multiple biopharmaceutical firms, it is not possible to objectively assert what these results signify. It is essential for such an index to be widely adopted by trial teams, investigators, and regulators to allow for more widespread comparison and validation across time and geographies. Even beyond the broader implications of representativeness, the metric can be more proximally useful to development of clinical trials by operationally demonstrating trial enrollment dynamics; for instance, the bimodal distribution observed for asthma trials (Fig. 2) may help understand how sites in asthma trials in the UK may have variable performance in terms of enrolling representative cohorts.

R-index can be readily computed for different populations and settings depending on local data availability. There is utility from a regulatory and scientific standpoint in understanding how two similar trials, such as the Covid-19 vaccine trials used in this study,

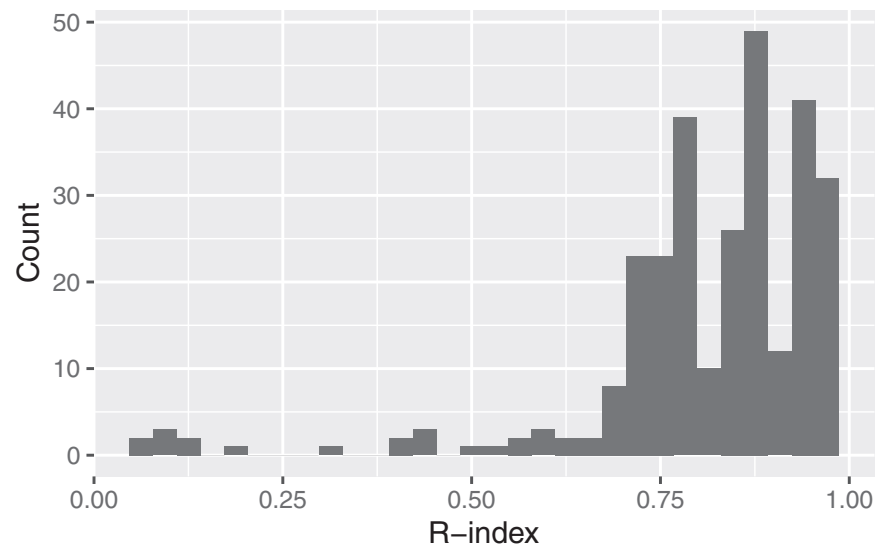


Fig. 1: Histogram showing distribution of R-index for study enrollment compared to local catchment populations across study site and trial combinations in the UK.

differ in terms of representativeness. Calculating R-index in this setting would be suitable for regulatory offices in different countries with some epidemiological and demographic expertise. It may also be possible to develop parametric statistical tests to assess differences between two R-index distributions to compare representativeness of two trials in the same setting or to test a null hypothesis that an R-index is not 1.0 or not statistically representative. R-index may obscure small but important error terms for small population groups. The

practical solution for investigators is to aim for an R of 1, as then even smaller groups will have 0 error. Even with this consideration, R-index remains a useful composite index to track the concept of representativeness in different settings, both historically and contemporaneously. Future studies may explore development of methods for accounting for such considerations.

There are moral, scientific, and regulatory imperatives to ensure that biopharmaceutical research is representative. While investigators should prioritize

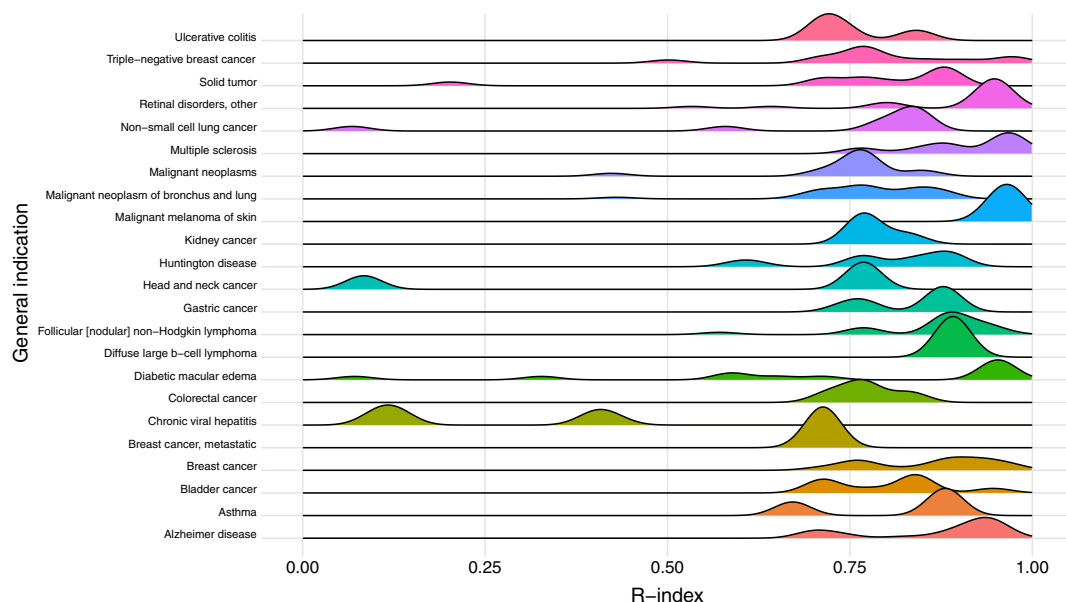


Fig. 2: Distribution of R-index values across trial sites by disease area in Roche UK trials. Each observation is a trial-site combination.

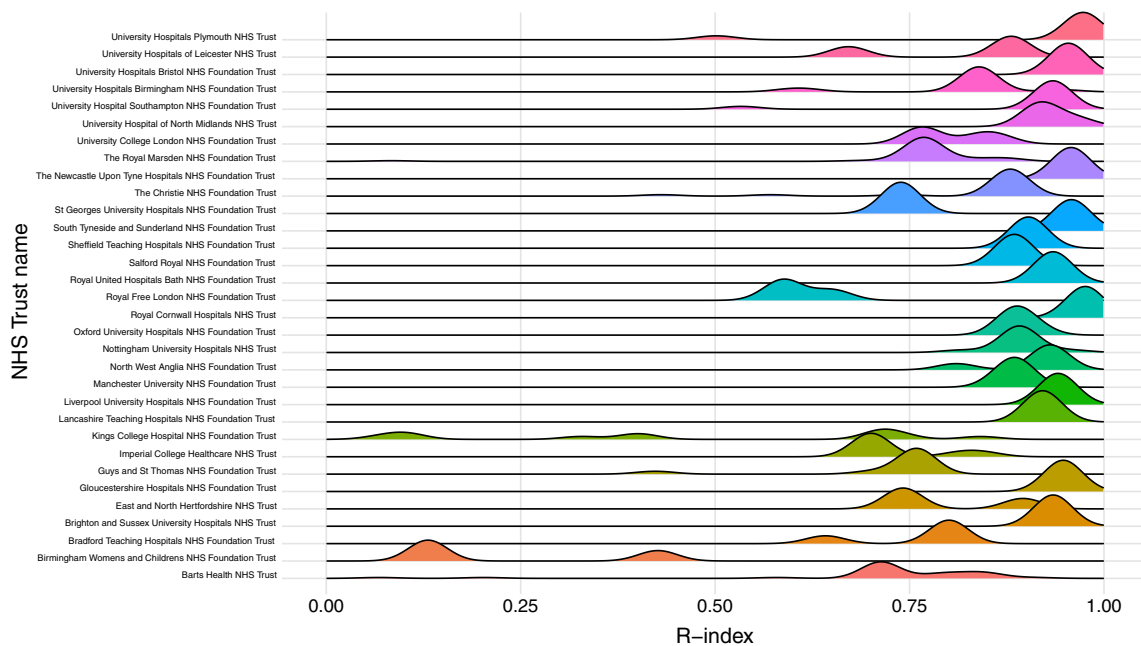


Fig. 3: Distribution of R-index values across trials and disease areas by NHS Trust in Roche UK trials. NHS, National Health Service.

representativeness for these reasons alone, R-index could be associated with better outcomes in longer-term, real-world studies, which can reveal that an intervention has lower effectiveness in a real-world setting than in a clinical trial. The difference between efficacy and effectiveness in these settings should be smaller for interventions with a higher R-index due to some

assurance that a trial's enrollment population is similar to the population that may be treated. While real-world effectiveness can decrease relative to trial efficacy for other reasons, it can be hypothesized that the R-index would be a robust, simple, univariate measure to help predict these longer-term outcomes using transportability analyses used in health economics and

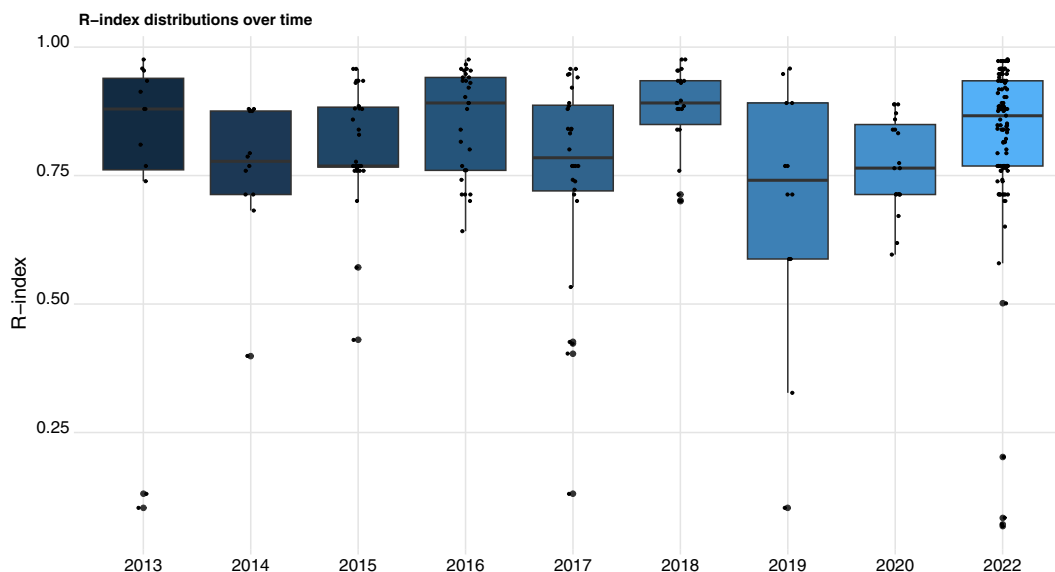


Fig. 4: Distribution of R-index values across trials and NHS sites across time, computed by midpoint year of trial location. The size of each data point is proportional to the sample size of the trial site. NHS, National Health Service.

Group	Trial, n	SEER, n	Trial, proportion	SEER, proportion
American Indian or Alaska Native	3	1384	0.35%	0.34%
Asian and Hawaiian/Pacific Islander	184	20,730	21.65%	5.09%
Black or African American	16	44,453	1.88%	10.92%
White	598	339,139	70.35%	83.31%
Other, multiple, unknown	49	1382	5.76%	0.34%
Sum	850	407,088	100%	100%

NSCLC, non-small cell lung cancer; SEER, Surveillance, Epidemiology, and End Results.

Table 2: Trial enrollment and SEER data for NSCLC.

statistics research.⁵ This could have considerable impact on the ability of modern therapeutics to benefit a global population and be advantageous to biopharmaceutical firms in predicting real-world effectiveness relative to trial efficacy.

Implementing the R-index should be approachable in any biopharmaceutical research or similar setting where representativeness is paramount. The examples in this study used publicly available census data on demographics. More advanced applications may require more careful assignment of specific populations into subsets. For example, a trial focused on metastatic NSCLC with specific inclusion and exclusion criteria may reflect a nonrepresentative sample of the US population, and an investigator may aim to estimate the R-index using published or measured disease burden data.

This study has several limitations. It is recognized that ontologies are limited in adequately categorizing ethnically diverse groups, and this could be a driver of nonrepresentative research. While this paper focused on single- or cross-study measurements using the same population ontologies, further work is needed to investigate methods for comparing R-indices in populations that use different ontologies. Because data were

unavailable at the NHS Trust level, this analysis did not incorporate local ethnicity-specific disease burden populations. This limitation manifests the assumption that disease risk is uniform across ethnic groups, which is generally not true. Therefore, this example only demonstrates how the R-index could be used. Caution should be taken in interpreting these results since a specific NHS Trust may have less representative enrollment than is ideal; however, at a national level, it could still be possible to have better representativeness at the trial level since most trials include enrollment from ≥ 1 NHS Trust. Trial sites could have nonrepresentative enrollment, while an overall trial may still be representative of a country or geography as a whole. Nevertheless, it can be argued that every trial site should aim to have representative enrollment and if trial site placement itself is representative, then these effects should yield representativeness at the level of the overarching geography. In looking at the R-index distributions, occasionally there are a few low outliers; these appear to be due to a relatively small number of data points such as observed in head and neck cancer, which has three R values available. Another limitation was that post hoc aggregates were developed to convert the tabulations to comparable groups. However, this is not unique to the R-index and would be a limitation in any error assessment in this setting, including conventional error metrics. It is also recognized that some studies may deliberately oversample patients by design to allow power for subgroup analysis in which, for example, there may be concern for differential safety or efficacy based on other research including early-stage studies. In this case, the R-index can still be calculated but may result in a lower value with all else being equal due to study design, although in theory the additional sample size for some subgroups could be accounted for in the overall formulation of R-index such that there is still assurance of representativeness accounting for the study design.

The goal of this study was to introduce the R-index, demonstrate its application, and advocate for broader assessment of metrics and methods in order to understand how they can improve patient outcomes. With the well-characterized and increasingly important need for inclusive research and population-representative data, the R-index and similar novel benchmarking parameters are critical measures for all clinical research, although further work is needed across institutions, sponsors, regulators, and payers to understand and agree upon standards for using the R-index as well as other methods and metrics relevant in this space. Given the complex considerations of characterizing human diversity in structured data, a summary metric such as the R-index overcomes the limitations of relying on absolute and relative error metrics alone. R-index is a temporally and geographically appropriate standardized metric for benchmarking and summarizing representativeness; it

	BNT162b2 trial	Population	Error	mRNA-1273 trial	Population	Error
White	0.829	0.616	0.213	0.792	0.616	0.176
Black or African American	0.093	0.124	0.031	0.102	0.124	0.022
Asian	0.043	0.060	0.017	0.046	0.060	0.014
Native American or Alaska Native	0.005	0.011	0.006	0.008	0.011	0.003
Native Hawaiian or Other Pacific Islander	0.002	0.002	0.000	0.002	0.002	0.000
Multiracial	0.023	0.186	0.163	0.042	0.186	0.144
Not reported	0.006	0.001	0.005	0.009	0.001	0.008

Absolute error was computed as the absolute difference between the trial and populations.

Table 3: Proportions of race and ethnicity in the overall BNT162b2 trial and mRNA-1273 trial (combined treatment and control arms) and corresponding estimates from 2020 US Census data.

exemplifies the type of novel approaches that must be developed in clinical research to ensure that the next era of transformational medicines is equitably developed and distributed.

Contributors

Spencer L James, MD: conceptualization, formal analysis, and writing.
Max Bourgognon, PhD: conceptualization, visualization, and methodology.

Patricia Pinto Vieira, BSc: conceptualization.

Bruno Jolain, MD: writing.

Sarah Bentouati, PharmD: visualization, methodology, and writing.

Emma Kipps, PhD: methodology and writing.

Kate Gillespie, PhD: methodology and writing.

Assaf Oron, PhD: methodology and writing.

Ruma Bhagat, MD: conceptualization.

Altovise Ewing, PhD: conceptualization.

Shalini Hede, PharmD: conceptualization and writing.

Keith Dawson, DNP: conceptualization.

Nicole Richie, PhD: conceptualization and writing.

The corresponding author (SLJ) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. Transparency: The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained. SLJ and EK had direct access and verified the underlying data reported in the manuscript.

Data sharing statement

Qualified researchers may request access to individual patient-level data through the clinical study data request platform (<https://vivli.org/>). Further details on Roche's criteria for eligible studies are available here (<https://vivli.org/members/ourmembers/>). For further details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here (https://www.roche.com/research_and_development/who_we_are_how_we_work/clinical_trials/our_commitment_to_data_sharing.htm).

Declaration of interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/and-declare:

Spencer L James, MD: employee of Genentech/Roche and owns Roche equity at the time of the study.

Max Bourgognon, PhD: none.

Patricia Pinto Vieira, BSc: employee of Roche Products Ltd and owns Roche equity at the time of the study.

Bruno Jolain, MD: none.

Sarah Bentouati, PharmD: none.

Emma Kipps, PhD: consulting fees from Novartis and Roche; honoraria and speaker fees from Pfizer and Novartis; speaker fees from AstraZeneca; support for meetings from Novartis.

Catherine Gillespie, PhD: none.

Assaf Oron, PhD: employee of the University of Washington.

Ruma Bhagat, MD: employee of Genentech/Roche and owns Roche equity at the time of the study.

Altovise Ewing, PhD: employee of Genentech/Roche and owns Roche equity at the time of the study.

Shalini Hede, PharmD: employee of Genentech/Roche.

Keith Dawson, DNP: employee of and has received support for attending meetings and/or travel from Genentech/Roche and owns Roche equity at the time of the study.

Nicole Richie, PhD: employee of Genentech/Roche and owns Roche equity at the time of the study.

Acknowledgements

This study was funded by F. Hoffmann-La Roche Ltd/Genentech, Inc., which contributed to the study design, data interpretation, and writing of this report. The authors thank Denise Kenski, PhD, of Nucleus Global for providing editorial assistance, which was funded by Genentech, Inc., in accordance with Good Publication Practice (GPP2022) guidelines. The authors also thank Johnny Wharton for critical insights.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclinm.2025.103079>.

References

- Oyer RA, Hurley P, Boehmer L, et al. Increasing racial and ethnic diversity in cancer clinical trials: an American Society of Clinical Oncology and Association of Community Cancer Centers joint research statement. *J Clin Oncol*. 2022;40:2163–2171. <https://doi.org/10.1200/JCO.22.00754>.
- Green AK, Trivedi N, Hsu JJ, et al. Despite the FDA's five-year plan, Black patients remain inadequately represented in clinical trials for drugs. *Health Aff*. 2022;41:368–374. <https://doi.org/10.1377/hlthaff.2021.01432>.
- US Food and Drug Administration. Diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials: guidance for industry [online]. <https://www.fda.gov/media/157635/download>; 2022. Accessed September 22, 2023.
- International Federation of Pharmaceutical Manufacturers and Associations. Diversity and inclusion in clinical trials: bioethical perspective and principles [online]. <https://www.ifpma.org/resource-centre/diversity-and-inclusion-in-clinical-trials-bioethical-perspective-and-principles/>; 2022. Accessed September 22, 2023.
- Degtiar I, Rose S. A review of generalizability and transportability. *Annu Rev Stat Appl*. 2023;10:501–524. <https://doi.org/10.1146/annurev-statistics-042522-103837>.
- US Census Bureau. 2020 Census results [online]. <https://www.census.gov/programs-surveys/decennial-census/decade/2020/2020-census-results.html>; 2023. Accessed September 22, 2023.
- Office for National Statistics. Census 2021 [online]. <https://census.gov.uk/>. Accessed September 22, 2023.
- Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med*. 2020;383:2603–2615. <https://doi.org/10.1056/NEJMoa2034577>.
- Baden LR, El Sahly HM, Essink B, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med*. 2021;384:403–416. <https://doi.org/10.1056/NEJMoa2035389>.
- Socinski MA, Jotte RM, Cappuzzo F, et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med*. 2018;378:2288–2301. <https://doi.org/10.1056/NEJMoa1716948>.
- Office for Health Improvement and Disparities. NHS acute (hospital) trust catchment populations. 2022 Rebate experimental statistics. <https://app.powerbi.com/view?r=eyJrJoiODZmNGQ0YzltZDAwZi00MzFiLWE4NzAtMzVmNTUwMTMThmMTVlIiwidCI6ImVlNGUxNDk5LTRhMzUtNGlyZS1hZDQ3LTVMmM2NmOWRlODY2NlslmMiOjh9>. Accessed November 18, 2024.
- UK Government. Ethnicity facts and figures. List of ethnic groups. 2011 Census [online]. <https://www.ethnicity-facts-figures.service.gov.uk/style-guide/ethnic-groups/#2011-census>. Accessed November 18, 2024.
- Murray CJ, Lozano R, Flaxman AD, et al. Robust metrics for assessing the performance of different verbal autopsy cause assignment methods in validation studies. *Popul Health Metr*. 2011;9:28. <https://doi.org/10.1186/1478-7954-9-28>.