



## OPEN ACCESS

## EDITED BY

Ming Hsien Tsai,  
Shin Kong Wu Ho-Su Memorial Hospital,  
Taiwan

## REVIEWED BY

Ziran Li,  
University of California, San Francisco,  
United States  
Nelson Lu,  
United States Food and Drug Administration,  
United States

## \*CORRESPONDENCE

Ralf J. Ludwig,  
✉ ralf.ludwig@uksh.de  
Uazman Alam,  
✉ uazman.alam@liverpool.ac.uk

RECEIVED 24 October 2024

ACCEPTED 13 February 2025

PUBLISHED 10 March 2025

## CITATION

Ludwig RJ, Anson M, Zirpel H, Thaci D,  
Olbrich H, Bieber K, Kridin K, Dempfle A,  
Curman P, Zhao SS and Alam U (2025) A  
comprehensive review of methodologies and  
application to use the real-world data and  
analytics platform TriNetX.  
*Front. Pharmacol.* 16:1516126.  
doi: 10.3389/fphar.2025.1516126

## COPYRIGHT

© 2025 Ludwig, Anson, Zirpel, Thaci, Olbrich,  
Bieber, Kridin, Dempfle, Curman, Zhao and  
Alam. 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.

# A comprehensive review of methodologies and application to use the real-world data and analytics platform TriNetX

Ralf J. Ludwig<sup>1,2,3\*</sup>, Matthew Anson<sup>4,5</sup>, Henner Zirpel<sup>3</sup>,  
Diamant Thaci<sup>3</sup>, Henning Olbrich<sup>2</sup>, Katja Bieber<sup>1</sup>,  
Khalaf Kridin<sup>1,6,7</sup>, Astrid Dempfle<sup>8</sup>, Philip Curman<sup>1,9,10,11</sup>,  
Sizheng S. Zhao<sup>12</sup> and Uazman Alam<sup>4,5,13\*</sup>

<sup>1</sup>Lübeck Institute of Experimental Dermatology, University of Lübeck, Lübeck, Germany, <sup>2</sup>Department of Dermatology, University Hospital Schleswig-Holstein Lübeck, Lübeck, Germany, <sup>3</sup>Institute and Comprehensive Centre for Inflammation Medicine, University-Hospital Schleswig-Holstein, Lübeck, Germany, <sup>4</sup>Department of Cardiovascular and Metabolic Medicine, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, United Kingdom, <sup>5</sup>Department of Medicine, Aintree University Hospital, Liverpool University Hospitals NHS Foundation Trust, Liverpool, United Kingdom, <sup>6</sup>Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel, <sup>7</sup>Unit of Dermatology and Skin Research Laboratory, Galilee Medical Center, Nahariya, Israel, <sup>8</sup>Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany, <sup>9</sup>Dermato-Venereology Clinic, Karolinska University Hospital, Stockholm, Sweden, <sup>10</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, <sup>11</sup>Dermatology and Venereology Division, Department of Medicine (Solna), Karolinska Institutet, Stockholm, Sweden, <sup>12</sup>Centre for Musculoskeletal Research at University of Manchester, Manchester, United Kingdom, <sup>13</sup>Centre for Biomechanics and Rehabilitation Technologies, Staffordshire University, Stoke-on-Trent, United Kingdom

Randomized controlled trials (RCTs) are the gold standard for evaluating the efficacy and safety of both pharmacological and non-pharmacological interventions. However, while they are designed to control confounders and ensure internal validity, their usually stringent inclusion and exclusion criteria often limit the generalizability of findings to broader patient populations. Moreover, RCTs are resource-intensive, frequently underpowered to detect rare adverse events, and sometimes narrowly focused due to their highly controlled environments. In contrast, real-world data (RWD), typically derived from electronic health records (EHRs) and claims databases, offers a valuable counterpart for answering research questions that may be impractical to address through RCTs. Recognizing this, the US Food and Drug Administration (FDA) has increasingly relied on real-world evidence (RWE) from RWD to support regulatory decisions and post-market surveillance. Platforms like TriNetX, that leverage large-scale RWD, facilitate collaborations between academia, industry, and healthcare organizations, and constitute an in-depth tool for retrieval and analysis of RWD. TriNetX's federated network architecture allows real-time, privacy-compliant data access, significantly enhancing the ability to conduct retrospective studies and refine clinical trial designs. With access to currently over 150 million EHRs, TriNetX has proven particularly effective in filling gaps left by RCTs, especially in the context of rare diseases, rare endpoints, and diverse patient populations. As the role of RWD in healthcare continues to expand, TriNetX stands out as a critical tool that complements traditional clinical trials, bridging the gap between controlled research environments and real-world

practice. This review provides a comprehensive analysis of the methodologies and applications of the TriNetX platform, highlighting its potential contribution to advance patient care and outcomes.

#### KEYWORDS

real-world data, TriNetX, cohort study, Kaplan–Meier estimator, drug discovery

## Introduction

Randomized controlled trials (RCTs) are the reference standard trial design for demonstrating causality of a specific intervention, pharmacological or non-pharmacological, which includes evaluation of efficacy and adverse events (Hariton and Locascio, 2018). Usually, stringent inclusion and exclusion criteria, as well as randomization, are applied, in part for safety, but also to offset the effects of both known and unknown confounders, whilst follow-up visits and timing of visits are highly controlled (Bhide et al., 2018). These stringent criteria minimize selection bias which may impact on the translatability (Hernán et al., 2004). However, RCTs are not feasible for all research questions and as such have a number of limitations, namely: 1. Lack of feasibility of high resource-intensive RCT for long term outcomes, 2. RCTs are often underpowered to detect harm, 3. Results of some RCTs may lack broad applicability due to their narrow eligibility criteria, relative small sample size, highly controlled implementation of interventions and comparators, and shorter duration of study. Often the focus or primary endpoint is on short-term surrogates or composite outcomes, which again may lack translatability in a real-world setting, especially for long-term outcomes. Additionally, in RCTs it is extremely challenging to address ethically challenging issues, such as the use of drugs in pregnancy or other vulnerable populations. Based on these considerations some proponents of real-world evidence (RWE) often highlight that RCTs may not be feasible, ethical, or timely (Hernán and Robins, 2016).

Unlike traditional clinical trial data, real-world data (RWD) are often generated from alternative sources often collected such as claims databases, patient generated data and most often electronic health records (EHRs). Consensus on the precise definition of RWD is lacking, however the most consistent definition is ‘data collected in a non-randomized controlled trial setting’ (Makady et al., 2017). The Food and Drug Administration (FDA), defines RWD as ‘the data relating to patient health status and/or the delivery of healthcare routinely collected from a variety of sources’ (U.S. Food and Drug Administration, 2024b). RWD is becoming increasingly important in healthcare decisions, where RCT data are lacking such as in rare diseases and off-label drug therapy, including use of drugs in vulnerable patient populations, such as during pregnancy. More recently in 2018, the FDA created a framework for the evaluation the potential use of RWE to support the approval of a new indication(s) for a drug already approved under the Federal Food, Drug, and Cosmetic Act and to help support post-approval drug study requirements. Importantly, RWD harbors significant potential for generation of RWE but also notably for designing and conducting subsequent confirmatory studies including RCTs. RWE through RWD analysis may answer questions that may not be otherwise addressed (Liu and Panagiotakos, 2022). Furthermore, RWD can serve as an

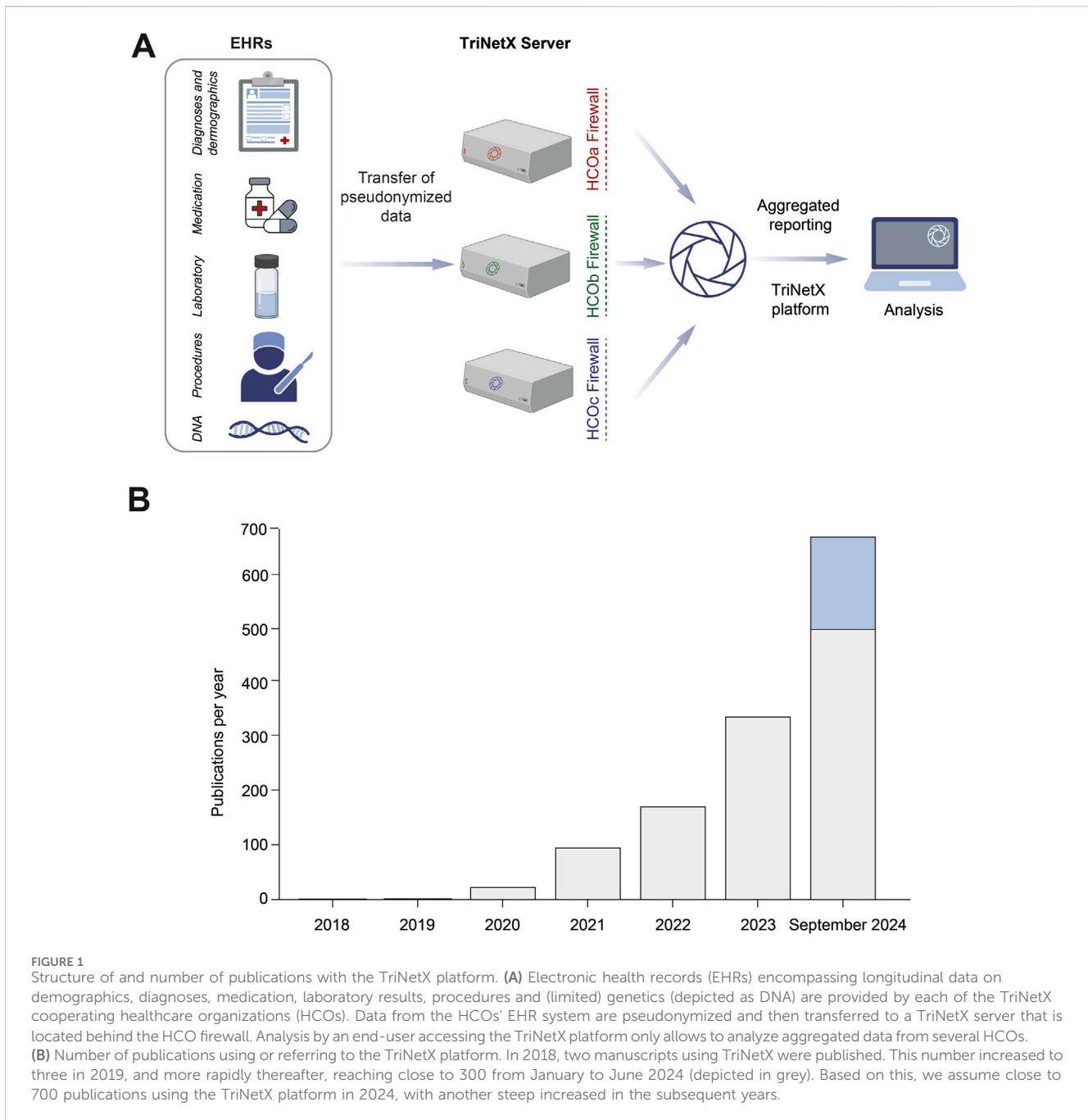
alternative to clinical cohort studies for assessing potential risk factors that cannot be randomized, such as patient characteristics (e.g., comorbidities, blood type, ethnicity) or lifestyle factors that are either ethically (e.g., smoking) or practically (e.g., long-term diet) unfeasible to randomize. This approach is particularly relevant for studying population risks, which conceptually cannot be examined in RCTs.

TriNetX (TriNetX LLC, Cambridge, MA, USA) is an extensive growing global network expanding from 55 healthcare organizations (HCOs) and 7 countries in 2017 (Palchuk et al., 2023) to 137 HCOs and 17 countries as of August 2024. It is often described as a real-world ecosystem given its real time continually updated nature. The platform is a federated network which facilitates collaboration between industry including pharmaceutical companies and research organizations, academia, HCOs and community-based HCOs. TriNetX has a security and governance model that provides a federated platform of RWD (electronic health records, datasets) which is Health Insurance Portability and Accountability Act (HIPAA), General Data Protection Regulation (GDPR), and Lei Geral de Proteção de Dados (LGPD)-compliant.

The aim of this paper is to review the methodologies, applications and published studies using the real-world data and analytics platform TriNetX.

## Setting the stage: A brief introduction to the TriNetX platform

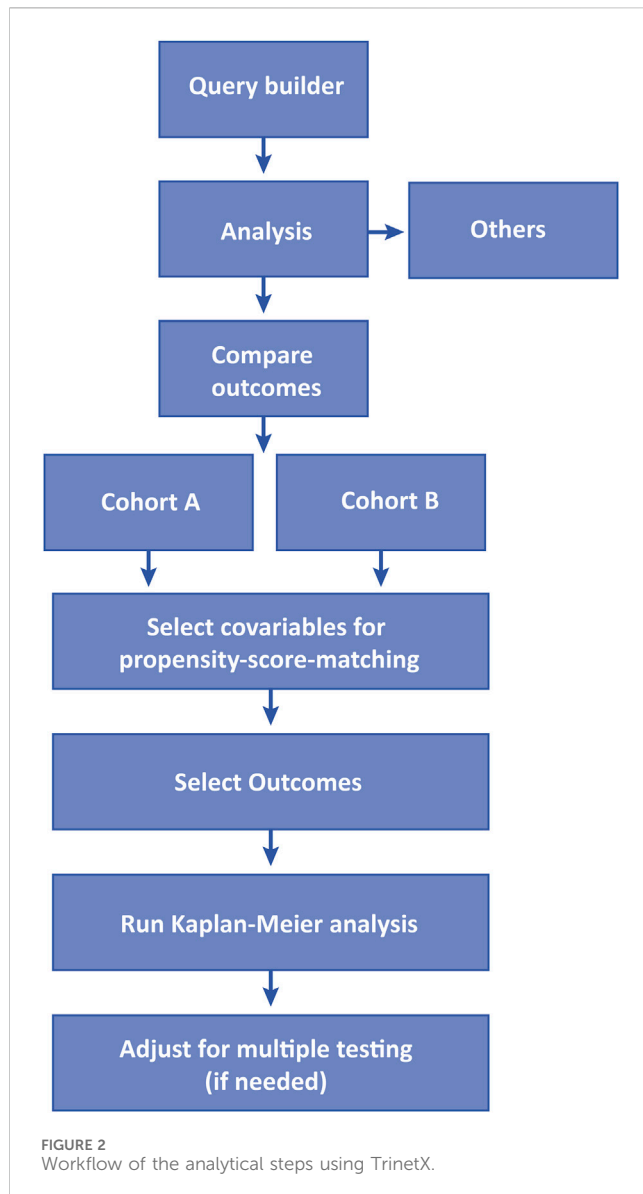
Given the challenges of sustainability of academia governed and financed clinical data repositories, encompassing granular electronic health records (EHRs), the private company TriNetX has successfully implemented an alternative approach for the funding and operational management of clinical data repositories in 2014. Here, industry partnerships with healthcare organizations (HCOs) led to the establishment of a large federated clinical data repository. In this setting, TriNetX operates the platform (including the legal framework and analytical tools), funding is provided by industry, and HCOs provide EHRs (Palchuk et al., 2023). From this cooperation, industry and academia can access over 150 million EHRs as of September 2024. Industry participants usually use the federated TriNetX database to query patient counts to challenge feasibility of in- and exclusion criteria for clinical trials. These patient counts can be queried globally, restricted to certain regions, or to single HCOs. This enables real-time iteration and the optimization of trial design and patient recruitment that has recently been described and reviewed in-depth elsewhere (Palchuk et al., 2023; Topaloglu and Palchuk, 2018; Claerhout et al., 2019; Campion et al., 2024). Conversely, the focus of this review is on the methodologies and application of the TriNetX network for epidemiological research. Access to the TriNetX is either by



commercial licensing agreement (industry) or by collaboration agreement (academia).

As depicted in [Figure 1A](#), at each HCO pseudonymized EHR data are transferred from the own clinical data management system to a server that is still behind the HCOs' firewall. If a query is made outside the HCO, only aggregated data from several HCOs can be retrieved. Thus, the retrieved datasets only contain de-identified data as per the de-identification standard defined in the Health Insurance Portability and Accountability Act (US federal law) Privacy Rule ([TriNetX. Publication Guidelines, 2017](#)). As such, studies with TriNetX typically do not require evaluation by an Ethical Review Board. Within the TriNetX network, EHRs can be retrieved and analyzed from five Collaborative Networks: The

Global Collaborative Network (as of September 2024, 153.5 million EHRs from 128 HCOs), the US Collaborative Network (117.2 million EHRs from 66 HCOs), the Latin America (LATAM) Collaborative Network (31.5 million EHRs from 32 HCOs), the Europe, Middle East, and Africa (EMEA) Collaborative Network (19.2 million EHRs from 26 HCOs), and the Asian Pacific (APAC) Collaborative Network (5.1 million EHRs from 14 HCOs). In addition, data from the HCO that one is affiliated with can be retrieved. The US and Global Collaborative Networks can be used with or without natural language processing (NLP). NLP allows to use data from unstructured documents to extract standardized information, such as medications and diagnoses. The Collaborative Networks are, with the exception of the Global



Collaborative Network, mostly distinct and do not contain overlapping data. Since the establishment of this infrastructure in 2014, research on the platform has rapidly increased: In 2018, two articles were published that were authored by TriNetX employees. The article by Dr. Stapff compared cardiovascular outcomes in patients treated with either sodium-glucose co-transporter-2 (SGLT2) inhibitors or dipeptidyl peptidase 4 inhibitors and demonstrated lower risks in those treated with SGLT2 inhibitors (Stapff, 2018). The article by Drs. Topaloglu and Palchuk reviewed the use of the TriNetX network to optimize clinical trial operations (Topaloglu and Palchuk, 2018). Since then, the number of publications using the TriNetX network has dramatically increased. In total, as of September 2024, there close to 1,000 publications, with 457 of these published between January and September 2024 (Figure 1B). These include publications in high-ranking journals such as Nature Medicine (Wang et al., 2024a), Lancet Rheumatology (Singla et al., 2023), and Lancet Psychiatry (Taquet et al., 2022). Overall, this indicates the high relevance of real-world data for biomedical research.

## In more detail: Study design, data retrieval and analysis on the TriNetX platform

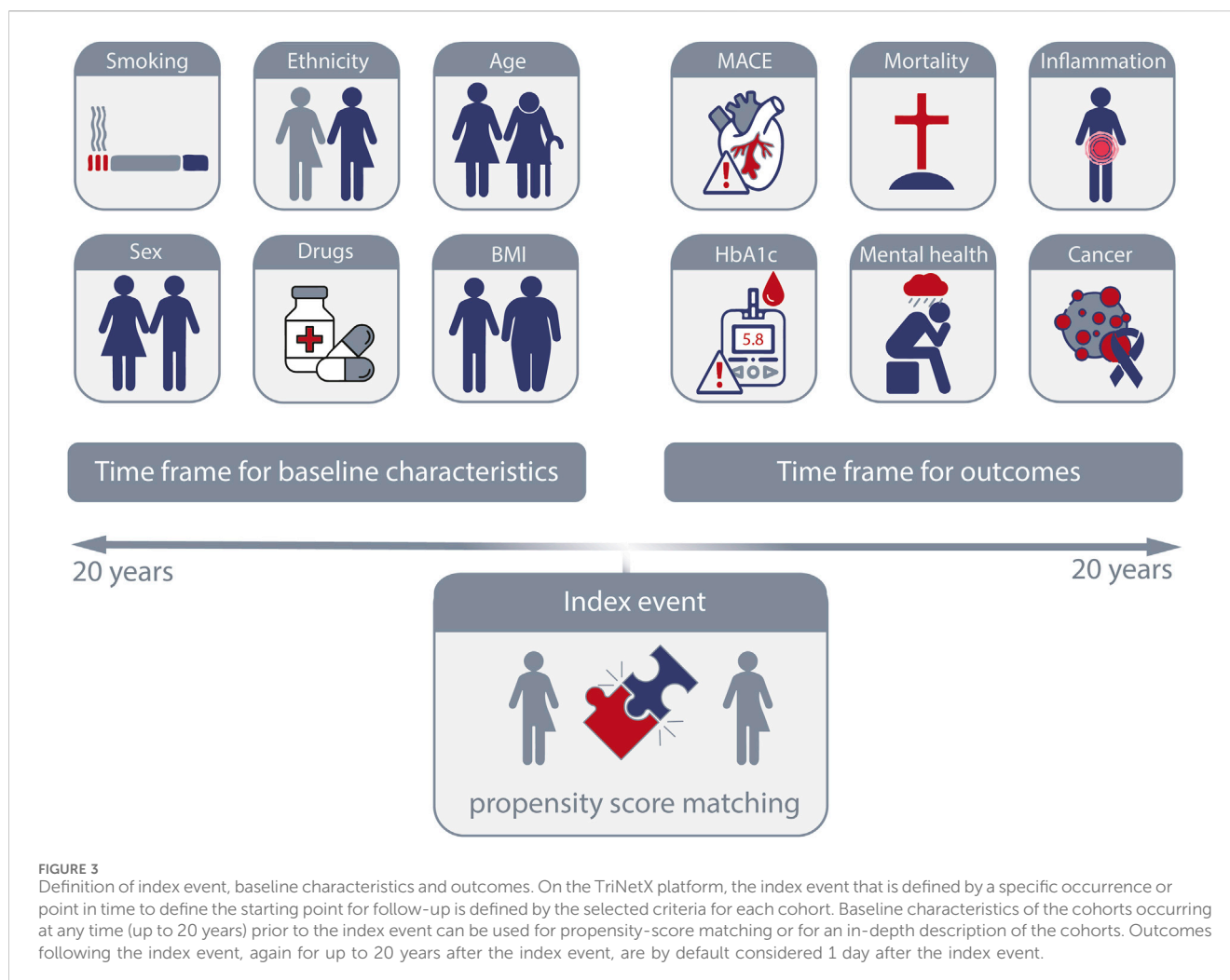
The TriNetX platform enables users to conduct retrospective studies by accessing deidentified EHRs. The retrospective study design and analysis tool are user-friendly, guiding users step-by-step to build a cohort and analyze outcomes. Initially, users create individual cohorts using a query builder, and in the subsequent step, these cohorts are analyzed for outcomes of interest (Figure 2).

### Query Builder

To initiate an analysis, users select the relevant Collaborative Network (Global, US, Latin America, “Europe, Middle East, and Africa”, or Asia Pacific) from which EHR data will be retrieved. Cohort construction then involves applying specific inclusion (“must have”) and exclusion (“cannot have”) criteria, with at least one inclusion criterion originating from the “diagnosis” category. Criteria are grouped according to international coding systems, including “demographics,” “diagnosis” (ICD-10-CM, ICD-9), “oncology” (ICD-O, AJCC, NAACCR, TNX curated), “procedures” (ICD-10-PCS, CPT, HCPCS, SNOMED), “medications” (VA, ATC), “labs” (LOINC), “genomics” (Gene), and “visits”. When multiple criteria are applied, they can be linked using the Boolean operators “and” (requiring all selected criteria to be present in the same individual) or “or” (requiring the individual to meet at least one of the criteria). Each criterion can be further refined with filters, such as specifying an age range for a diagnosis or a route of administration for a medication. Additional details can be integrated into each group, including the number of occurrences, unique instances, creating related groups, or adding temporal constraints to define relationships between groups. Multiple groups can be used to define patient cohorts, and these groups can also be connected using the Boolean operators “and” or “or”. This approach is particularly useful for selecting patients with a specific diagnosis within a certain timeframe after an initial procedure, or for including individuals with predefined baseline characteristics. Time constraints can also serve as an alternative to the built-in tool to exclude outcomes before the index event (referring to a specific occurrence and/or point in time used to define the starting point for follow-up in the cohort). If more than one group of definitions is used to define a cohort, the index event must be chosen from among them. Once the cohort is defined, users can explore all cohort characteristics using the “explore cohort” function, which includes demographics (e.g., age, sex, ethnicity), diagnoses, oncology, procedures, medications, labs, and genomics. These characteristics encompass all data reported by healthcare organizations (HCOs) at any given time. The platform also provides additional tools following the query builder, including “analyze criteria,” “rate of arrival,” “summary statistics,” and “analysis,” although the latter three tools may be of limited use for most academic users.

### Analytics

Once the study cohort(s) have been constructed, by default, the analytics tool currently offers six distinct analysis options: “Analyze Outcomes”, “Compare Outcomes”, “Compare Cohorts”, “Treatment Pathways”, “Incidence and Prevalence”, and “Advanced Explore Cohort”. There are also additional analytical



tools under development that may be accessible for beta-testing. Among these, “Compare Outcomes” is the most commonly used function for research purposes. While other options such as “Treatment Pathways” and “Advanced Explore Cohort” may also provide valuable insights, their applicability can be limited by data availability and ease of use. Similar considerations apply to the “Treatment Pathway Analysis” that has only reported in few publications (Ghosh et al., 2024; Henao-Martínez et al., 2023). Therefore, the focus is often on the “Compare Outcomes” function, though the general methodology can be extended to the other functions.

In the “Compare Outcomes” function, users select two predefined cohorts from the query builder to compare. The platform then guides users through a structured five-step analysis process, organized around three time segments: Index event, time prior to the index event, and the period following the index event (Figure 3). For each cohort, an index event is defined, such as the first diagnosis of a disease or the initial treatment with a specific drug after diagnosis. This index event corresponds to the baseline in prospective studies. The three time segments include the period before the index event (up to 20 years), the index event itself (on the same day), and the period following the index event (up to 20 years).

After defining the index event and timeframe for the first segment, both cohorts are evaluated for differences in characteristics (present before the index event) that could serve as potential confounders for the outcomes of interest. These key characteristics are then used as criteria for propensity score matching at the index event. The criteria for matching typically include “Demographics”, “Diagnoses”, “Oncology”, “Procedures”, “Medications”, “Labs”, and “Genomics.” Baseline characteristics are analyzed before matching using the “Baseline Comparison Statistics” function, which checks for differences between the unmatched cohorts and ensures that the matching criteria are appropriate. In the “Balance Cohorts” step, users finalize the propensity score matching by selecting input variables and reviewing the matching results. In the final segment, two analysis functions are available: “Explore Outcomes” and “Outcomes”. The “Explore Outcomes” function serves as a preliminary screening tool, identifying statistical differences between cohorts without performing matching. Outcomes marked during this step are further analyzed in the “Outcomes” function, where users refine the analysis by setting the timeframe after the index event (e.g., 1 day to 3 months or 1–5 years) and confirming the index event. The outcome analysis can be conducted using one of four available methods: “Measures of

Association” (risk ratio, risk difference, odds ratio with 95% confidence intervals, t-test), “Kaplan-Meier Analysis” (log-rank test, hazard ratio), “Number of Instances” (t-test), and “Lab Results” (t-test for numeric values, chi-square test for categorical values). As a recent addition, follow-up time for each cohort can also be retrieved.

For the first three analysis methods, it is recommended to exclude patients who experienced the outcome event before the index event. This exclusion can be enabled in the “Options” section. The final results are provided in the “Results” section, and the complete analysis, including network details, inclusion and exclusion criteria, matching characteristics before and after matching, analysis methods, and individual outcome analyses, can be downloaded.

## Statistical considerations

### Sample size and power analysis

In prospective studies, sample size calculation should be performed before data collection to ensure adequate power (“as many as necessary”) while keeping cost and potential risk to participants low (“as few as possible”). For retrospective studies, sample size may be limited by the availability of data and an *a priori* analysis of statistical power should ensure that a planned investigation is adequately powered to detect meaningful differences or associations (Kim and Seo, 2013). If such minimal power requirements are fulfilled, all available data should be used in retrospective studies and there is no need to decrease sample size to only obtain, e.g., 80% power. To perform a power calculation, the specific test (e.g., log rank test, chi-square test, t-test) and the anticipated effect sizes (e.g., hazard ratio, proportions, standardized mean difference) as well as the desired significance level need to be specified. For this, free online power calculators and dedicated programs are available, such as G\*Power (Heinrich Heine Universität Düsseldorf, 2020) or in R, the “powerSurvEpi” package may be used to calculate the power for survival analysis based on the Cox proportional hazards model (CRAN.R-project, 2021). Generally, the very large sample size available in the TriNetX platform offers high power even for rare exposures (index events) and for rare outcomes. A key advantage of utilizing all available data in retrospective studies is the ability to generate more precise effect estimates. However, concerns regarding “overpowering” may arise. Overpowering, however, mostly relates to prospective study designs where excessive sample sizes may lead to detecting statistically significant but clinically trivial differences. In retrospective analyses, the interpretation of results should focus on clinical significance, ensuring that findings are not only statistically meaningful but also relevant to patient outcomes and clinical decision-making.

### Propensity score matching (PSM)

To alleviate confounding and allow for a better baseline comparability of groups, PSM should be used. For PSM, the TriNetX platform utilizes the user-defined set of covariates for each patient within the cohorts. Using logistic regression, the probability that a patient belongs to a particular cohort is calculated. This predicted probability, known as the propensity

score, reflects the probability that a subject will receive an exposure based on their covariate profile. Once propensity scores are generated, the system tries to match each patient from the smaller cohort with a corresponding patient from the larger cohort who has a sufficiently similar propensity score. This matching process results in a subset of both cohorts, where the distribution of covariates is balanced between the groups, and aims for less confounded comparison and analysis of outcomes between the two cohorts. This method is widely used in epidemiological and clinical research to reduce bias in observational studies, where randomization is not feasible. PSM may also be advantageous over conventional multiple Cox proportional hazards analysis (Elze et al., 2017) if there is a large number of potential confounders. If performed on the platform, the number of covariates for PSM is limited based on sample size due to computational restrictions. Hence, for PSM considering many covariates, sample sizes need to be reduced. This can be achieved by adding time windows, limiting the age range, or adding additional terms, e.g., exclusion of deceased patients.

### Correction for multiple testing

Correcting the significance level for multiple testing in epidemiological research aims to control the family wise error rate, reducing the likelihood of type I errors - incorrectly rejecting a true null hypothesis (i.e., finding a false positive). In large-scale epidemiological analyses the chance of detecting false positives increases as the number of tests increases. Multiple testing corrections, such as the Bonferroni correction or less conservative alternatives such as the false discovery rate (FDR) correction, potentially ensure that any significant findings are more likely to reflect true associations rather than random chance. In contrast, Dr. Rothman advocates that no adjustments for multiple comparisons are necessary in epidemiological research (Rothman, 1990). While adjusting for multiple comparisons reduces type I errors, it simultaneously increases the risk of type II errors, leading to potentially missed true associations. In Dr. Rothman’s view, the underlying “universal null hypothesis,” which attributes findings to chance, contradicts the empirical research principle that nature operates under discoverable laws. By avoiding such adjustments, researchers may achieve more accurate interpretations of real-world data and feel more empowered to explore potentially significant findings without the constraints of overly conservative corrections (Rothman, 1990). Ultimately, the decision of how stringently, i.e., selecting Bonferroni or FDR, to adjust for multiple comparisons in epidemiological research should be guided by the specific context and objectives of the study. While corrections can prevent false positives, they may also obscure true associations, especially in exploratory research. Therefore, researchers should weigh the trade-offs and consider the balance between rigor and discovery in their analytical approach.

### Odds (OR) or hazard ratio (HR)

Analyzing RWD, both odds ratios (OR) and hazard ratios (HR) can be used to compare the likelihood of events between two groups. However, they differ fundamentally in their consideration of time. OR provide a static measure, comparing the likelihood of an event occurring between two groups without taking temporal information into account. This renders OR suitable for cross-sectional and case-

control studies where timing is not a factor. In contrast, hazard ratios incorporate the element of time, making them ideal for survival analysis or time-to-event studies where censoring is present, i.e., not every person experiences the event during the available follow-up time. HRs compare the rate at which events occur over time between groups and can establish whether the event occurs significantly earlier in one group compared to the other.

### Considering the proportionality assumption

If HR are used, the assumption of proportional hazards must be met, meaning that the hazard rates between groups follow the same trend over time, differing by a constant factor (the HR). If this assumption is violated, the HR can be misleading. The TriNetX platform incorporates a test for proportionality, that was developed by Grambsch and Therneau (Grambsch and Therneau, 1994), which relies on the scaled Schoenfeld residuals. The Schoenfeld residual for a parameter reflects the difference between the observed value for a patient and the expected value based on the model, adjusted for variability through scaling by the variance. Interpreting the resulting chi-square and p-value require careful consideration, as a non-significant p-value does not prove that the assumption is fulfilled while at the same time even small deviations from this assumption (which may not relevantly affect the model) may lead to nominally significant p-values due to the large sample size. TriNetX advises users to treat these metrics as qualitative guides rather than exact quantitative measures (TriNetX, 2024). Specifically, users are encouraged not to view a significance level of 0.05 as a strict cutoff for decision-making. Instead, chi-square values should be interpreted broadly, with larger values indicating relevant deviations from proportionality while small values suggest that the proportionality assumption may hold. If there is evidence that the proportionality assumption is seriously violated, an analysis based on ORs can still be performed.

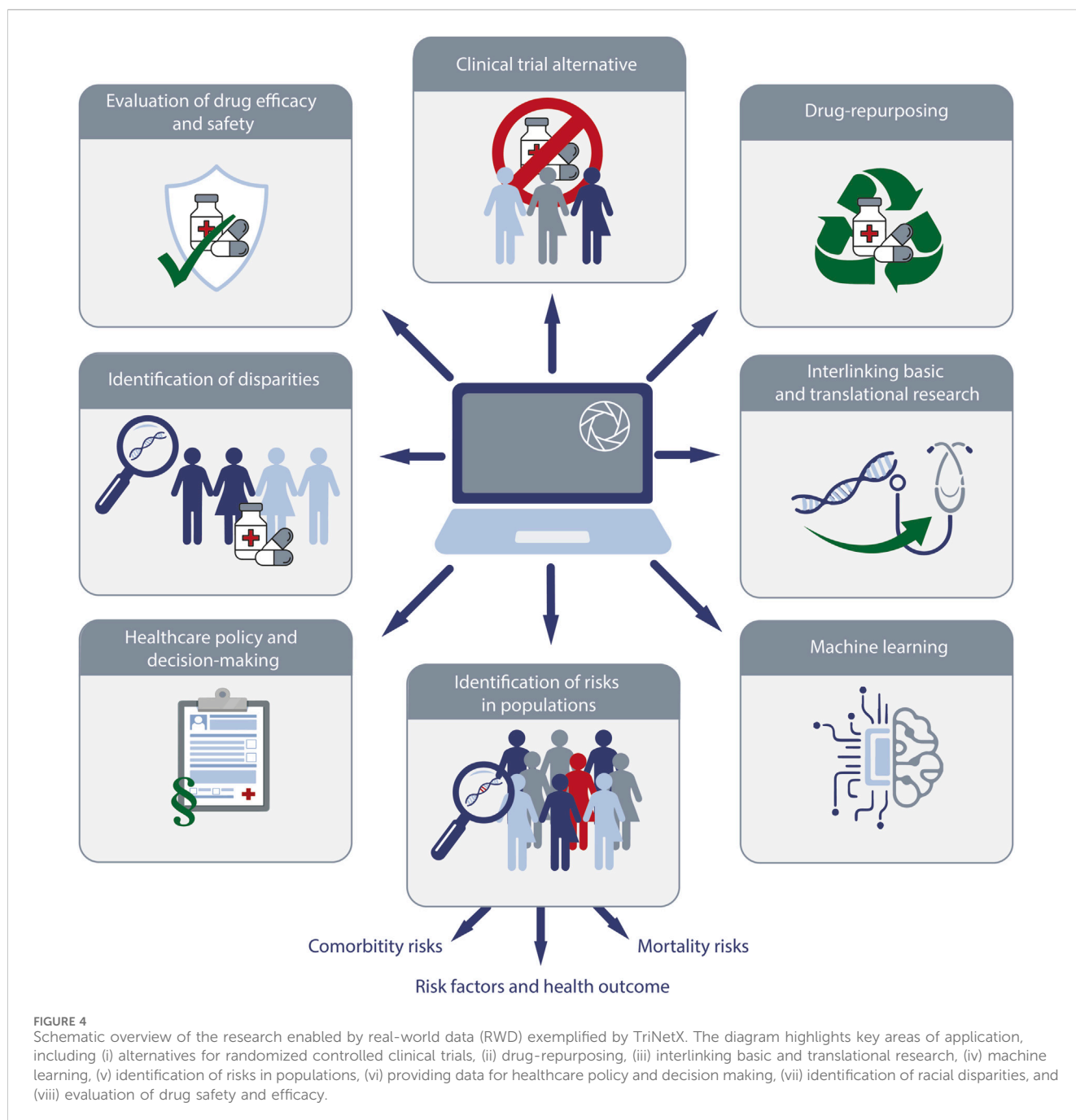
### Importance of sensitivity analyses

To mitigate potential bias and confounding based on the selection of covariates for PSM and the outcomes, it is recommended to perform several sensitivity analyses by conducting the same analysis across different cohort variations. This helps to assess the robustness of the findings and reduces the risk that results are unduly influenced by specific cohort characteristics or assumptions. Sensitivity analyses can provide greater confidence in the validity of the study's conclusions by testing the consistency of outcomes under different conditions. Examples for cohort variations are.

- Excluding time periods shortly after index event helps mitigate the risk of including outcomes that may not be directly related to the exposure or intervention. Early outcomes might be due to existing conditions rather than the event or intervention under study. The length of such time periods to exclude should be based on subject matter knowledge on the minimum time for the event of interest to develop. By excluding these time periods, the analysis focuses on outcomes more likely attributable to the exposure.
- Including a minimal baseline time window of patient records before inclusion in the study better ensures that there is adequate data to assess the patient's condition before the index event. This allows for better control of confounding factors and ensures that the baseline characteristics are well-defined, improving the accuracy of the propensity score matching and subsequent analyses.
- Use of alternative covariates for PSM may result in varying levels of balance between groups, and thus allows to determine whether findings are robust across different scenarios or if they are sensitive to the specific variables selected. This ensures that the matching process captures a wide range of relevant factors, minimizing bias and enhancing the credibility of the results. It also provides a more comprehensive assessment of the potential confounders, making the findings more reliable and generalizable to different contexts.
- Excluding outcomes occurring before the index event at data retrieval instead of after PSM (which is the default option in TriNetX) better mitigates bias from influencing the matching process. If outcomes are excluded after PSM, the matching may be biased by the presence of those outcomes. Excluding them earlier in the process ensures that the matching is based purely on baseline covariates, leading to more valid comparisons between cohorts. Taking this into account, excluding outcomes before the index event prior to PSM may be better suited as a primary analysis, and the default option in TriNetX as a sensitivity analysis.
- Modifying the strictness of cohort definitions allows to explore the impact of varying levels of participant selection on study outcomes. A more lenient definition might include a broader population, enhancing generalizability but potentially introducing more variability or distortion. Conversely, a stricter definition can focus on a more specific population, improving internal validity but possibly limiting the applicability of the findings to broader groups. Selecting this for a sensitivity analysis helps determine whether the study's conclusions are robust across different definitions or if they are highly dependent on specific criteria.
- Similar considerations apply when changing outcome definitions (more lenient or stricter).
- Changing the time window of data retrieval enables evaluation of the impact of different time frames on study outcomes. A wider time window might capture more events, potentially increasing statistical power but also introducing variability from long-term changes in treatment practices or patient characteristics. Conversely, a narrower time window focuses on more recent or immediate outcomes, which may reduce noise but limit the number of events available for analysis.

### Identification of risks in populations

The use of real-world data platforms like TriNetX has significantly enhanced our ability to identify and analyze risks within various populations (Figure 4). This involves assessing risk factors, comorbidities, and mortality across diverse demographic and clinical groups. The ability to access a vast dataset allows for more comprehensive studies, leading to better understanding and management of health risks.



## Risk factors and health outcomes

Understanding the myriad risk factors that contribute to health outcomes is crucial for effective disease prevention and management. The TriNetX platform, with its extensive RWD, enables researchers to identify and analyze these risk factors across diverse populations, uncovering critical insights that inform clinical practice and public health strategies. One example is the identification of pruritus (chronic itching) as a risk factor for severe health conditions. Studies have shown that pruritus can be associated with an increased risk of hematological malignancy and autoimmune diseases (Deng et al., 2022; Raap et al., 2023; Caldas et al., 2023). These findings underscore the importance of recognizing seemingly minor symptoms as potential indicators of more severe underlying health issues.

## Comorbidity risk

Assessing comorbidities within patient populations is crucial for understanding the interplay between different health conditions and their combined impact on patient outcomes. The TriNetX database facilitates extensive analysis of comorbid conditions, offering valuable insights into how these conditions coexist and influence each other. For example, a retrospective cohort study using TriNetX data revealed an increased risk of autoimmune diseases in patients who had recovered from COVID-19, highlighting the long-term health implications of the virus (Chang et al., 2023). Similarly, another study examined long-term cardiovascular outcomes in COVID-19 survivors, finding significant risks associated with the virus in non-vaccinated populations (Wang et al., 2022). Other



examples of studies focused on comorbid health conditions utilizing the TriNetX platform include the investigation into the co-existence of various autoimmune disorders among patients with pre-existing autoimmune blistering diseases (Kasperkiewicz et al., 2023) and comorbid risk of cardiovascular outcomes in patients with systemic lupus erythematosus (Olbrich et al., 2023) or prurigo nodularis (Olbrich et al., 2024). These studies emphasize the importance of monitoring patients for comorbid conditions to ensure comprehensive healthcare management.

### Mortality risk

Understanding mortality risk is a key component of population health management. The TriNetX platform's extensive dataset enables researchers to conduct in-depth analyses of factors contributing to mortality, aiding in the development of strategies to mitigate these risks. For example, research utilizing TriNetX demonstrated that statin use is associated with a reduced risk of liver disease, hepatocellular carcinoma, and liver-related mortality (Vell et al., 2023). Similarly, another study comparing treatments for acute stroke, such as tenecteplase *versus* alteplase, have identified significant differences in mortality and bleeding complications, informing clinical decision-making (Murphy et al., 2023). All-cause mortality in patients with type 2 diabetes undergoing different treatments (Riley et al., 2023) and the mortality risk in Stevens-Johnsons syndrome and toxic epidermal necrolysis managed in different ways (Ozhathil et al., 2024) have been investigated using the TriNetX platform. Matched case-control studies for mortality among different conditions can also be performed (Boch et al., 2023). These types of investigations into different aspects of mortality risk provide critical insights into the nature of different diseases and the effectiveness of various treatments and their impact on patient survival.

### Where clinical trial data are lacking

We have previously detailed that RWD can provide RWE where RCT data are lacking, e.g., investigation of rare (adverse) events, head-to-head comparisons, or off-label medications, or, in particular, where RCTs are unlikely to occur do to cost implications. Recently, Anson et al. published a RWD study utilizing the TriNetX platform of multimodal pharmacotherapy of a combination of metformin, pioglitazone, sodium-glucose-linked-cotransporter-2 inhibitor (SGLT2i), and glucagon-like peptide-1 receptor agonist (GLP-1RA) *versus* a more conventional glucocentric treatment (insulin and sulphonylureas) for cardiovascular outcomes in the treatment of type 2 diabetes (Anson et al., 2024a). The pathophysiology of type 2 diabetes is complex including  $\beta$  cell failure, muscle, and liver insulin resistance, increased renal glucose reabsorption, reduced incretin response, a cell insulin resistance, reduced secretion of and increased resistance to GLP-1 and glucose-dependent insulinotropic polypeptide (GIP) and appetite dysregulation (DeFronzo, 2009). As such, this paper assessed the real-world evidence in which combination therapy targeted these pathophysiological mechanisms. It demonstrated that intensive combination therapy, targeting the distinct pathophysiological defects in type 2 diabetes was associated with favorable cardiorenal outcomes, compared to conventional insulin and sulphonylurea therapy. Additionally, a potent GLP-1 RA, namely, semaglutide was reassessed for a link to suicidal ideation,

however a large TriNetX Network analysis (240,618 patients) did not demonstrate higher risks of suicidal ideation with semaglutide compared with non-GLP1R agonist anti-obesity or anti-diabetes medications (Wang et al., 2024a). Importantly previous RCTs have been too small and of too short duration to effectively answer this important question necessitating RWD to generate RWE was required to provide reassurance to patients and clinicians to the widely used pharmacotherapy. Similarly in the dermatology space, Kridin et al. demonstrated the putative negative effect of oral corticosteroid vs. topical corticosteroid for bullous pemphigoid on mortality at 1 and 3 years (Kridin et al., 2024a). Such cardiovascular and mortality outcome studies are unlikely to ever be undertaken in an RCT setting but are essential to capture the long-term impact of treatment outcomes.

More recently, the specific methodology of target trial emulation has been used to apply the principles of RCTs to observational data (Matthews et al., 2022). This approach can help estimate the causal effect of an intervention and has several key steps which include; 1. Designing a hypothetical RCT, 2. Specify the protocol including eligibility criteria, treatment strategies, outcomes, etc., and 3. Utilizing the available observational data to emulate the hypothetical RCT. Such methodology can be utilized with the TriNetX platform. Recently, Wang et al. (Wang et al., 2024b), demonstrated that semaglutide was associated with lower risks for tobacco used disorder-related healthcare measures in patients with comorbid T2DM and tobacco use disorder compared with other anti-diabetes medications including other GLP-1RAs. This study utilized target trial unlauded approach which specifically improves the quality of observational studies investigating the effects of interventions by providing a structured framework to help overcome the limitations of observational studies and improve causal inference.

### Evaluation of drug effectiveness and safety

TriNetX data can also serve as a resource for evaluating both the effectiveness and safety profiles of pharmaceuticals currently employed in clinical practice. Given the absence of disease severity documentation within the network, surrogate markers of effectiveness must be relied upon. While possible exceptions exist, such as the assessment of vaccination efficacy using the corresponding infectious disease as an outcome, this approach has not, to our knowledge, been utilized with EHR data from TriNetX. Additionally, known sequelae of a given disease, such as psoriatic arthritis in relation to psoriasis, or mortality risks, can serve as indicators of drug efficacy (Singla et al., 2023). To evaluate drug efficacy or safety, the study should retrieve EHRs with an ICD10-code for the disease of interest, e.g., acne vulgaris. Patients with this condition are then allocated to distinct groups based on their exposure to one of two possible treatments. Whenever feasible, treatments should be chosen based on their position within the treatment hierarchy, allowing for comparison of outcomes between two same-line treatments. Otherwise, the risk of "confounding-by-indication" (in this context failure of first line treatment) might be high and this cannot be controlled by the propensity score matching described above as the relevant characteristic (previous first line treatment) will mostly be present in one group. TriNetX has been

widely used to study drug efficacy and safety. Examples are the safety of isotretinoin use in acne vulgaris and biological treatment of chronic inflammatory diseases, that are described in more detail below.

### Safety of isotretinoin in acne vulgaris treatment

Acne is a chronic inflammatory disease of the pilosebaceous unit, driven by immune responses affecting adolescents and young adults. Clinically, it manifests as both non-inflammatory lesions, i.e., comedones, as well as inflammatory lesions, such as papules, pustules, nodules, and cysts. Predilection sites are the face, chest, and back (Moradi Tuchayi et al., 2015). Oral antibiotics, especially tetracyclines are commonly used first line treatments for severe acne (Costa et al., 2021). Another first line treatment of severe acne is oral isotretinoin (Reynolds et al., 2024). However, the use of isotretinoin in patients with acne has been associated with potentially serious risks, including the potential development of psychiatric disorders and inflammatory bowel disease (Chen et al., 2022; Taylor et al., 2023). Yet, the risks of psychiatric disorders and inflammatory bowel disease in patients with acne exposed to oral isotretinoin have remained a matter of debate (Wright et al., 2021; Tan et al., 2024). With TriNetX, large cohorts of in patients with acne exposed to isotretinoin or antibiotics can be retrieved. This allows the investigation of a substantial number of EHRs along with extensive propensity-score matching for potential confounding factors for the risk to develop psychiatric disorders or inflammatory bowel disease. Regarding psychiatric disorders, a recent study using data from TriNetX, demonstrated that isotretinoin is associated with a lower risk of depression, post-traumatic stress disorder, anxiety, bipolar disorder, schizophrenia, and adjustment disorder compared to oral antibiotics. The risk of major depressive disorder and suicidal attempts were similar between the groups, although the risk of suicidal ideation was higher with isotretinoin (Kridin and Ludwig, 2023a). This finding has recently been replicated in a meta-analysis of 24 studies (Tan et al., 2024), underscoring the validity of the results obtained from TriNetX. Risk of Crohn's disease or ulcerative colitis (UC), in patients with acne exposed to either oral isotretinoin or oral antibiotics was explored in a recent study which reported no increased lifetime risk of CD or UC in the isotretinoin group compared to the antibiotics group. In a time-restricted analysis of the same dataset, a transient increase in UC risk during the first 6 months of isotretinoin use was observed, with five additional cases per 10,000 patients, but this risk diminished when later time points were investigated. Interestingly, the risk of irritable bowel syndrome, that was used as an additional endpoint, was lower in patients exposed to isotretinoin (Kridin and Ludwig, 2024; Kridin and Ludwig, 2023b).

### Biological treatment of chronic inflammatory diseases

Biological treatment has become an invaluable treatment for many non-communicable, chronic inflammatory diseases and several malignancies (U.S. Food and Drug Administration, 2024a). Data from clinical trials relating to these compounds, however, span a limited time frame. Hence, additional benefits or previously unforeseen adverse events, e.g., reactivation of tuberculosis following TNF inhibition (Solovic et al., 2010), are

likely to be overlooked. In this context, RWD can substantially enhance the assessment of drug safety and effectiveness. One landmark paper in this context has been published on the risk of inflammatory arthritis in patients with psoriasis treated with different classes of biologics (Singla et al., 2023). Psoriasis is a chronic inflammatory disease primarily affecting the skin (Griffiths et al., 2021) that is associated with a high extracutaneous comorbidity, such as depression, psoriatic arthritis, metabolic disorders and cardiovascular disease, of which psoriatic arthritis is most frequent (Gershater et al., 2024). The aforementioned study compared the risk of developing psoriatic or inflammatory arthritis in psoriasis patients who were treated with tumor necrosis factor inhibitors (TNFi), interleukin-17 inhibitors (IL-17i), or IL-12/23i. Here, patients treated with IL-12/23i had a significantly lower risk of developing psoriatic or inflammatory arthritis compared to those treated with TNF inhibitors. Given these findings are validated in prospective studies or data from prospective registries, this would have a significant impact on the management of psoriasis (Singla et al., 2023). Other studies in psoriasis using TriNetX contrasted the risk of infections (Kridin et al., 2023a), malignancies (Kridin et al., 2024b), and cardiovascular diseases (Kridin et al., 2023b) in psoriasis patients treated with different groups of biologics. In line with the findings regarding the risk of psoriatic or inflammatory arthritis, the risk of infections was lower in patients treated with either IL17-or IL12/23-inhibitors, as opposed to those managed by TNFi (Kridin et al., 2023a). Similar findings were made regarding the risk of malignancies (Kridin et al., 2024b). Contrasting the risk of cardiovascular diseases in psoriasis patients treated with either IL17-or IL12/23 inhibitors found no significant differences (Kridin et al., 2023b).

### Predicting treatment outcomes

The inhibition of immune checkpoints, which are negative regulators of immune activation that normally limit antitumor responses, has led to almost unparalleled clinical responses in patients with various types of cancer (Ribas and Wolchok, 2018). However, immune checkpoint inhibition is potentially complicated by (cutaneous) immune-related adverse events. These are often mild to moderate, are treatable and reversible. In some cases, immune-related adverse events can lead to lasting damage and death (Ramos-Casals et al., 2020). Cutaneous immune-related adverse events or presence of certain autoantibodies causing these, may indicate a positive response to immune checkpoint inhibition, and improved patient survival (Hasan Ali et al., 2020). To address this assumption in more detail, a retrospective cohort study used data from TriNetX, retrieving patients with cancer treated with immune checkpoint inhibition who developed cutaneous immune-related adverse events and matched with patients who did not. Development of cutaneous immune-related adverse events within 6 months after initiation of therapy was analyzed, adjusted for demographics, cancer type, and stage. Presence of cutaneous immune-related adverse, such as pruritus, drug eruption, xerosis, nonspecific rashes, psoriasis, and lichen planus/lichenoid dermatitis, were associated with a reduced mortality risk. The authors concluded that cutaneous immune-related adverse are strongly linked to better treatment responses and improved survival (Tang et al., 2022).

Taken together, these studies demonstrate that TriNetX is a useful tool for retrospectively analyzing drug effectiveness and

safety. It is important to note that the lack of established causality must be considered when interpreting this data, in particular potential confounding by factors not available for consideration in the propensity score matching. In the interim, however, these insights serve as a valuable resource for identifying potential risks and guiding treatment selection.

## Drug-repurposing

Investigation of drug effects in diseases outside their approved indications can accelerate and facilitate drug development for otherwise neglected diseases. Drug-repurposing strategies include *in silico* predictions on drug pathways or *in vitro* and animal investigations, but also observational studies. RWD can aid the identification of potential repurposing candidates, and several use cases have demonstrated beneficial effects of drugs previously unrelated to a certain disease. In the field of cancer research, TriNetX helped evaluate survival rates in patients with head and neck cancer and beneficial effects of metformin (Gaertner et al., 2024) and statins (Wüster et al., 2023) were found. By contrast, other drugs such as ibuprofen showed unfavorable results in this regard (Ebeling et al., 2024). These negative findings are, however, equally important as those that demonstrate beneficial effects. Specifically, the findings of a reduced survival of ibuprofen-exposed patients with head and neck cancer, compared to non-exposed patients urges caution when selecting pain medication in these patients. In patients with breast cancer, also, statins reduced the risk of all-cause death (Bucci et al., 2024). Further, several drug analyses supported by TriNetX were conducted for neuropsychiatric diseases given the orphan status and insufficient interventional options in many of them. For neurodegenerative diseases, reduced prevalence of Alzheimer's disease (Silva et al., 2023) and Parkinson's disease (Silva et al., 2024) was found in patients treated systemically with calcineurin inhibitors. Likewise, in a study which received significant public attention, a reduced risk of dementia was observed in persons exposed to the recombinant shingles vaccine (Taquet et al., 2024). Reduced frequencies of sleep and anxiety disorders were documented in patients treated with brain-penetrant calcium channel blockers (Colbourne and Harrison, 2022; Colbourne et al., 2021). Also, outcomes in several substance abuse disorders were significantly affected by drug-repurposing candidates: A drug screen identified eight medications that led to more frequent remissions in patients with opioid use, including psychiatric drugs as well as propranolol and prazosin (Patel et al., 2023). Cannabis use disorder and relapses were less frequent in patients treated with semaglutide (Wang et al., 2024c). Most notably, ketamine was identified as an efficient treatment for cocaine use disorder by first employing an AI-based screening strategy, expert review, and subsequent RWD confirmation in the TriNetX database (Gao et al., 2023a). The same prediction algorithm was also used to identify drugs that delay the progression of cataract in patients with diabetes mellitus (Gao et al., 2023b). Recently, a study comparing off-label tirzepatide use in people living with obesity but without T2D to those treated with semaglutide use reported a significant reduction in incident risk of developing T2D and serves as a vital hypothesis generating foundation for future definitive RCTs (Anson et al., 2024b). Taken together, TriNetX is a powerful tool to gather

retrospective evidence for the effects of repurposing candidate drugs, however, confirmatory clinical trials are necessary.

## Identification of ethnical disparities

RCTs may lack ethnic variation with the overrepresentation of White participants in particular (Kaakour et al., 2022). Although there have been efforts to recruit underrepresented minority ethnic groups, the lack of RCT data has resulted in a clear evidence gap in which drug effect may differ due to underlying ethnicity-specific biological mechanisms. In the NICE guidance for type 2 diabetes [NG28], a specific research recommendation was based on determining the 'effectiveness of SGLT2 inhibitors for different ethnic groups'. Taking this forwards, Anson et al. completed a real-world study of 98,094 participants propensity score matched (1:1). The subsequent analysis demonstrated that Black (regarding mortality and ischemic heart disease) and Asian (regarding heart failure and hospitalization) individuals demonstrate a greater degree of benefit compared to White individuals (Anson et al., 2023). When considering dermatological conditions, Kasperkiewicz et al., using data from the TriNetX platform, demonstrated that while the risk of mortality is comparable between White and skin of color (SOC) patients with immunobullous diseases, SOC patients have a significantly higher risk of hospitalization and post-inflammatory hyperpigmentation (Kasperkiewicz et al., 2025). Hypotheses generated were either higher rates of disease activity or inequality in access to healthcare. Clearly, these data need translating into studies of ethnicity-orientated disease mechanisms or healthcare access. Other RWD studies of dermatological conditions, namely, prurigo nodularis, a chronic inflammatory skin disorder which is intensely pruritic and has nodules of a hyperkeratotic nature on the torso and extremities, has also shown ethnic disparities in Black patients. There was greater mortality and possible systemic inflammation in Black patients (Sutaria et al., 2022). Although, these RWD demonstrate association rather than causality, they provide adequate data for the development of future RCTs.

## Provide data for decision-makers to set healthcare policies

RWD provides decision makers with a second pillar of evidence, including information on treatments, therapy pathways, and outcomes which can be stratified by baseline demographics and clinical characteristics. RWD can offer reassurance to clinicians and the wider public that evidence borne out in RCTs is (or sometimes is not), applicable to real world populations. RCTs have strict inclusion and discontinuation criteria, but the practice of clinical medicine is highly heterogenous, and many therapies can, and are prescribed to patients who do not necessarily fully satisfy the inclusion criteria of the study where the drug was initially investigated in. RWD includes data from a broader range of patients, capturing variations in age, gender and comorbidities; allowing decision makers to understand how interventions work across diverse groups, leading to more inclusive and effective policies. By analysing outcomes stratified by certain baseline characteristics, decision-makers can promote a personalised medicine approach to their populations, ultimately

improving clinical outcomes of patients. RCTs are limited by the lack of longer term follow up, and some beneficial and harmful outcomes of interventions are only captured in the extension phase of RCTs, which are often not done. RWD can follow patients over extended periods, often over decades, providing insights into the long-term safety and efficacy of treatments. RWD is often live, and continuously updated. The agility of such data ecosystems allows decision makers to react quickly to emerging health issues and novel therapies, subsequently allowing the timely adaptation of policy to respond to new threats. Finally, where the disease of interest is rare, or the intervention novel, or not widely used in a subset of the population, RWD can highlight areas where existing evidence is lacking, provide supplementary rationale to support novel RCTs, and guide future research investment, particularly in populations and therapies that may otherwise be overlooked. This use case of the TriNetX platform was especially useful during the COVID-19 pandemic, where RWD addressed critical questions within a short time-frame (Kovvuru et al., 2021; D'Silva et al., 2021; Taquet et al., 2021; Rivera-Caravaca et al., 2021).

## Linking machine learning and TriNetX

Within TriNetX, the LUCID feature enables researchers to utilize complex statistical tools and machine learning *in situ*, enabling the ability to download data while decreasing the time it takes to explore whether the data is fit for the research question proposed. RWD has been used to develop and validate a risk prediction model for pancreatic ductal adenocarcinoma. Using a large dataset of over 1.5 million cases and controls, 87 different features ranging from pre-existing diagnosis, medication and demographic data were harnessed. Authors were able to internally and externally validate their model; and simulate deployment of their model for early prediction of 6–18 months prior to diagnosis. The model captured over 3x more patients at risk of pancreatic ductal adenocarcinoma compared to existing screening programs of individuals at similar risk levels (Jia et al., 2023). Additionally, RWD has informed research evaluating different equations and methods of estimating glomerular filtration rate (GFR) in children, with further research required to confirm these findings which otherwise, without large populations to study, would go unasked, and ultimately, unanswered (Dagenais et al., 2024).

## Interlink with basic and translational research

Integrating RWD into larger research questions enables the foundation of basic and translational research to be firmly rooted in practical, everyday experiences. More specifically, RWD allows to generate hypotheses and validate mechanisms observed in basic research. Conversely, RWD may also be used to confirm findings from basic or translational research, thus providing a holistic understanding of biological processes across diverse settings. Ultimately, RWD serves as a vital bridge between fundamental research and real-world applications, enriching the overall research landscape. This interlink of basic and translational

research with RWD is an emerging field, with, so far, few publications in the field. As an example, using RWD to confirm and expand upon translational research findings, researchers investigated the molecular effects of anti-tumor necrosis factor (TNF) treatment on vaccination responses. This was prompted by recent findings questioning whether vaccination of patients with chronic, non-communicable inflammatory diseases under TNF-targeting treatment leads to impaired vaccine-induced immune responses and protection against breakthrough infections. The study used COVID-19 vaccination to investigate IgG subclass levels, IgG-Fc glycosylation patterns, B cell subsets, and antibody effector functions in patients on anti-TNF or other immunosuppressive treatments. Using TriNetX, the risk of SARS-CoV-2 breakthrough infections in these patients was assessed. Anti-TNF treatment reduced long-term levels of all anti-spike protein IgG subclasses with low galactosylation and sialylation. Initially, reactivated memory B cells produced highly galactosylated and sialylated IgG antibodies, which declined after each booster, especially in the elderly. Reduced IgG1 levels in these anti-TNF treated patients were linked to lower functional activity and a higher risk of COVID-19 (Buhre et al., 2025).

## Strengths and limitations of RWD

RWD has become increasingly valuable in healthcare, offering significant insights beyond traditional clinical trials. The volume of data captured by some ecosystems can exceed over 150 million individuals, enhancing the reliability of any findings described using such datasets. Data generated from RWD is more applicable to the lived experience of individuals and can be seen as 'RCT-affirming' studies, applicable to real world clinical practice. RWD often include the individual over many years, sometimes for most of their life. Such repositories of longitudinal data allow analysis of significantly large populations to track longer term health outcomes and treatment adherences, otherwise seldom investigated by other means, or if so, come at great expense and inconvenience. Research using even the largest of databases is highly cost-effective and more resource efficient than conducting new, large, multi-centre clinical trials. Additionally, novel treatment regimens can be analysed using RWD and be put forward as hypothesis generating, exploring associations that may have otherwise not gained the interest of sponsors or industry.

The limitations on utilizing RWD must be acknowledged. Firstly, by the very nature of the data source and lack of randomization, confounding variables, if unaccounted or unadjusted for, can bias any results generated using RWD. Importantly, causality cannot be established with observational data, only correlation. If RWD spans multiple countries, or multiple health institutions, variability in practice, coding for diagnosis and management of conditions can lead to high degrees of heterogeneity in data, and if stratification by location is not possible, may complicate interpretation and generalizability of findings. Besides variability in practice at the nation or healthcare system level, intra-caregiver variability can subject RWD to further inconsistencies, coding errors and missing data, further compounding the accuracy and reliability of such data sources. Incorporating and homogenizing unstructured data inputted from

various settings can be challenging, and lead to further inaccuracies. Although RWD incorporates a larger more diverse population pool compared to RCTs, for RWD that originate from economies with large health, and access to health inequalities, such as the United States, underserved and overlooked communities may still be underreported. To overcome many of the confounding variables as described, complex and sometimes technically advanced statistical analysis is required to balance cohorts, which may lead to data attrition. Additional limitations that apply to most RWD databases, including TriNetX, are lack of confidence in drug exposure, lack of documented disease severity and missing information on patient reported outcomes.

## Strengths and limitations of the TriNetX platform

As of September 2024, access to over 150 million EHRs is provided by TriNetX for academic research. This large, and continuously growing number of EHRs allows to investigate rare diseases such as epidermolysis bullosa acquisita (Kridin et al., 2023c) or hereditary angioedema (Sylvestre et al., 2021). Another advantage of the large dataset is the possibility to account for bias stemming from confounding factors, by use of extensive propensity score matching. Furthermore, this also allows to perform several sensitivity analyses, by changing the investigated time periods, adding minimal baseline time windows, generation of age-, sex-, or ethnicity-stratified groups, as well as changing definitions of index event and outcome definitions. From the TriNetX networks that can be accessed by default, this applies to the Global and the US Collaborative Network, as each of these encompass well over 150 million EHRs. The other Networks contain, in most cases, too few EHRs to allow investigation of rare diseases or an in-depth propensity score matching. In line with this notion, in the US Collaborative Network, the provided data is highly granular, i.e., demographics, diagnoses, medications and procedures are well coded. In other networks, e.g., the LATAM Collaborative Network some diagnoses seem underreported, exemplified by 0.02% EHRs indicating nicotine dependence within a total of over 20 million EHRs. Based on surveys, this proportion is expected to be around 11% (Andrade et al., 2002). Another example that relates to coding practices is the racial/ethnicity documentation in the EMEA Collaborative network. Here all EHRs are documented with an Unknown Ethnicity, and close to 85% are documented with an Unknown Race.

While the Global and US Collaborative Networks provide data for detailed research, other networks have limitations in certain areas. The high granularity of data required for specific analyses is primarily restricted to the US, as the Global Network is predominantly composed of US data. These gaps are not attributed to the TriNetX database but rather result from the legislation concerning data protection and coding practices in countries outside the US. Adopting data protection policies and documentation requirements globally would significantly improve the data quality within TriNetX. Such improvements would enable more comprehensive and detailed studies across diverse populations and regions. Addressing these documentation and policy issues could unlock the full potential of RWD within and outside the

TriNetX platform, establishing an unparalleled resource for global health research.

Within TriNetX a comprehensive and easy to use analytical pipeline is embedded. Amongst the seven analyses provided by default, the Compare Outcomes function is most commonly used. This and the other provided analyses are described in detail above (Setting the stage: Study design, data retrieval and analysis on the TriNetX platform). Overall, the navigation and running the analysis is very intuitive. At completion of the analysis all results can be downloaded. Depending on local legislation and access rules investigators may be able to download additional data. The default analyses available on the TriNetX platform effectively address the majority of research questions commonly explored. Indeed, most of the publications using the database, used this analytical tool. However, legislation concerning data protection also limit research with the TriNetX platform outside the US. More specifically, download of large datasets is not possible. This is, for example, needed to apply more complex analytical tools, e.g., AI, to determine predictors of the disease (Jia et al., 2023). Within the US it is also possible to build, and train own analytical models. This, for example, allows to cluster patients for safety and efficacy analyses, as well as more refined statistical analyses, e.g., use of the non-parametric test (Mann-Whitney) instead of t-test for laboratory values, which are often skewed. This feature would also enable to re-identify patients from the HCO of the researcher for closer monitoring and/or in-depth molecular analysis. A limitation that is related to the analytical pipeline itself is the missing option to build more complex outcomes. Currently, one or more outcomes can only be joined with an *or* denominator. It is currently not possible to implement time constraints or other more complex outcome definitions. For some diseases, especially those that cannot directly be coded with an ICD10-code alone, a more complex definition of outcomes would significantly add to the analytical pipeline embedded on the TriNetX platform.

If treatments are investigated on the platform, these must be considered as intention-to-treat analyses in most cases because drug compliance is not systematically captured on the platform. In more detail, to ensure that a given drug is prescribed for the investigated indication, a respective time-constrain has to be implemented when retrieving the data. This increases the likelihood that this drug was prescribed for the indication under investigation. However, compliance is not documented and can only be assumed using proxies. For example, by analyzing the number of events of drug prescription following the index event. Alternatively, a second (or third) documentation of drug prescription after a certain time period following the index event can increase the likelihood of continued drug exposure. If pseudonymized datasets are accessible (on a large-scale, currently only in the USA) this can be analyzed in more detail. However, drug compliance can only be documented when the same HCO documents the prescription.

In the past, TriNetX has continuously increased the number of EHRs available on the platform by expanding the network of contributing HCOs within existing networks. In addition, otherwise underrepresented areas are increasingly in the focus of TriNetX, exemplified by the rapidly growing LATAM Collaborative Network that mainly include EHRs from Brazil. Furthermore, the analytical platform has been extended. These additions to the analytical pipeline were, to a major part, driven by the request of

the researchers using the platform. Specifically, follow-up time is provided for most analyses since 2024, and the Cox Proportional Hazard Models has recently been added to the analytical pipeline. Several other analyses are under development, such as patient clustering, burden of illness and logistic regression.

Taken together, the TriNetX platform is an ideal instrument to address clinically highly relevant research questions. The main disadvantage relates to legislation and coding practices, rather than platform-inherent limitations. Therefore, changing legislation regarding data protection outside the US would significantly promote real-world-data research. As the TriNetX platform already is already among the leading platforms, research using TriNetX would benefit exponentially from this change in legislation.

## Conclusion

In conclusion, RCTs remain the cornerstone for establishing causality in clinical research. However, their limitations, including narrow eligibility criteria, resource intensity, and focus on short-term outcomes, restrict their applicability in broader, real-world settings. RWD not only complement RCTs but also extend beyond them by offering critical insights from more diverse populations and enabling the exploration of longer-term outcomes. Furthermore, RWD can address questions that RCTs are not equipped to answer, particularly in areas where traditional trials are impractical or unethical. This includes studying rare diseases, assessing off-label drug use, and evaluating interventions in populations often underrepresented in clinical trials. By capturing the complexities of real-world clinical practice, RWD provides a broader and more comprehensive perspective on treatment effectiveness and safety, ultimately filling the gaps left by RCTs. As the importance of RWD grows, further advancements in data governance, analytical tools, and global standardization will be crucial for maximizing its potential. Ultimately, the integration of RWD and RCTs holds the promise of more comprehensive and inclusive evidence generation, driving improvements in patient care and healthcare policy.

## Author contributions

RL: Conceptualization, Writing–original draft, Writing–review and editing. MA: Writing–original draft, Writing–review and editing. HZ: Writing–original draft, Writing–review and editing. DT: Writing–original draft, Writing–review and editing. HO: Writing–original draft, Writing–review and editing. KB: Visualization, Writing–original draft, Writing–review and editing. KK: Writing–original draft, Writing–review and editing. AD: Writing–original draft, Writing–review and editing. PC:

## References

Andrade, L., Walters, E. E., Gentil, V., and Laurenti, R. (2002). Prevalence of ICD-10 mental disorders in a catchment area in the city of São Paulo, Brazil. *Soc. Psychiatry Psychiatr. Epidemiol.* 37 (7), 316–325. doi:10.1007/s00127-002-0551-x

Writing–original draft, Writing–review and editing. SZ: Writing–original draft, Writing–review and editing. UA: Writing–original draft, Writing–review and editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Cluster of Excellence “Precision Medicine in Chronic Inflammation” (EXC 2167), the Collaborative Research Center “PANTAU” (SFB 1526), the individual research grant LU877 25/1, all from the Deutsche Forschungsgemeinschaft, the Schleswig-Holstein Excellence-Chair Program from the State of Schleswig-Holstein, Sinergia Unravel principles of self-organization in injured tissue (CRSII5\_202,301/1) from the Swiss National Science Foundation.

## Conflict of interest

HZ has received financial support for attending meetings and/or travel from Pfizer, UCB Pharma, Ammirall, Janssen, and TriNetX. HO has received financial support to attend meetings from TriNetX. RL has received honoraria for speaking or consulting or has obtained research grants from Novartis, Lilly, Bayer, Dompe, Synthon, Argen-X, Pharmaxis, CSL, TriNetX and Monasterium Laboratories during the last 3 years.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Generative AI was used for the creation of this manuscript. ChatGPT was used for editing.

## 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.

Anson, M., Henney, A. E., Broadwell, N., Zhao, S. S., Ibarburu, G. H., Lip, G. Y. H., et al. (2024b). Incidence of new onset type 2 diabetes in adults living with obesity treated with tirzepatide or semaglutide: real world evidence from an international retrospective cohort study. *eClinicalMedicine* 75, 102777. doi:10.1016/j.eclinm.2024.102777

- Anson, M., Malik, A., Zhao, S. S., Austin, P., Ibarburu, G. H., Jaffar, S., et al. (2024a). Treating type 2 diabetes with early, intensive, multimodal pharmacotherapy: real-world evidence from an international collaborative database. *J. Diabetes Res.* 2024, 3470654. doi:10.1155/2024/3470654
- Anson, M., Zhao, S. S., Essa, H., Austin, P., Ibarburu, G. H., Lip, G. Y. H., et al. (2023). Metformin and SGLT2i as first-line combination therapy in type 2 diabetes: a real-world study with a focus on ethnicity. *Clin. Ther.* 45 (12), 1259–1265. doi:10.1016/j.clinthera.2023.07.026
- Bhude, A., Shah, P. S., and Acharya, G. (2018). A simplified guide to randomized controlled trials. *Acta Obstet. Gynecol. Scand.* 97 (4), 380–387. doi:10.1111/aogs.13309
- Boch, K., Zirpel, H., Thaci, D., Mruwat, N., Zillikens, D., Ludwig, R. J., et al. (2023). Mortality in eight autoimmune bullous diseases: a global large-scale retrospective cohort study. *J. Eur. Acad. Dermatol. Venereol.* 37 (4), e535–e537. doi:10.1111/jdv.18700
- Bucci, T., Gue, Y., Dobson, R., Palmieri, C., Pignatelli, P., and Lip, G. Y. H. (2024). Statin use is associated with a lower risk of all-cause death in patients with breast cancer treated with anthracycline containing regimens: a global federated health database analysis. *Clin. Exp. Med.* 24 (1), 124. doi:10.1007/s10238-024-01395-z
- Buhre, J. S., Pongracz, T., Geisen, U. M., Schubert, M., Wang, W., Nouta, J., et al. (2025). Anti-TNF therapy impairs both short- and long-term IgG responses after repeated vaccination. *Allergy* 80 (2), 423–439. doi:10.1111/all.16241
- Caldas, S., Pagan, A. D., Correa Da Rosa, J., Kim, B. S., and Ungar, B. (2023). Irritable bowel syndrome and interstitial cystitis/bladder pain syndrome are associated with pruritus bidirectionally in U.S. adults. *Arch. Dermatol. Res.* 315 (9), 2745–2747. doi:10.1007/s00403-023-02684-5
- Campion, T. R., Sholle, E. T., Abedian, S., Fuld, X., McGregor, R., Lewis, A. N., et al. (2024). Implementation of a commercial federated network of electronic health record data to enable sponsor-initiated clinical trials at an academic medical center. *Int. J. Med. Inf.* 182, 105322. doi:10.1016/j.ijmedinf.2023.105322
- Chang, R., Yen-Ting Chen, T., Wang, S. I., Hung, Y. M., Chen, H. Y., and Wei, C. C. J. (2023). Risk of autoimmune diseases in patients with COVID-19: a retrospective cohort study. *EClinicalMedicine* 56, 101783. doi:10.1016/j.eclinm.2022.101783
- Chen, Y. H., Wang, W. M., Chung, C. H., Tsao, C. H., Chien, W. C., and Hung, C. T. (2022). Risk of psychiatric disorders in patients taking isotretinoin: a nationwide, population-based, cohort study in Taiwan. *J. Affect. Disord.* 296, 277–282. doi:10.1016/j.jad.2021.09.055
- Claerhout, B., Kalra, D., Mueller, C., Singh, G., Ammour, N., Meloni, L., et al. (2019). Federated electronic health records research technology to support clinical trial protocol optimization: evidence from EHR4CR and the InSite platform. *J. Biomed. Inf.* 90, 103090. doi:10.1016/j.jbi.2018.12.004
- Colbourne, L., and Harrison, P. J. (2022). Brain-penetrant calcium channel blockers are associated with a reduced incidence of neuropsychiatric disorders. *Mol. Psychiatry* 27 (9), 3904–3912. doi:10.1038/s41380-022-01615-6
- Colbourne, L., Luciano, S., and Harrison, P. J. (2021). Onset and recurrence of psychiatric disorders associated with anti-hypertensive drug classes. *Transl. Psychiatry* 11 (1), 319. doi:10.1038/s41398-021-01444-1
- Costa, C. S., Bagatin, E., Yang, Z., Pacheco, R. L., Magin, P., De Sá Urtiga Santos, L., et al. (2021). Systemic pharmacological treatments for acne: an overview of systematic reviews, editor. *Cochrane Database Syst. Rev.* (11). doi:10.1002/14651858.CD014917
- CRAN.R-project (2021). powerSurvEpi: power and sample size calculation for survival analysis of epidemiological studies. Available at: <https://cran.r-project.org/web/packages/powerSurvEpi/index.html>.
- Dagenais, S., Radi, Z., Burnham, J., Goodman, S., Bell, G., Corrigan, B., et al. (2024). Comparing five equations to estimate glomerular filtration rate or creatinine clearance and assign individuals to KDIGO categories across the full age spectrum using real world data. *Clin. Transl. Sci.* 17 (2), e13726. doi:10.1111/cts.13726
- DeFronzo, R. A. (2009). Banting Lecture. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. *Diabetes* 58 (4), 773–795. doi:10.2337/db09-9028
- Deng, J., Parthasarathy, V., Adawi, W., Bordeaux, Z., Sutaria, N., Gami, A., et al. (2022). Risk of hematologic cancer in patients with undifferentiated pruritus. *JAMA Dermatol* 158 (7), 791–795. doi:10.1001/jamadermatol.2022.1562
- D'Silva, K. M., Jorge, A., Cohen, A., McCormick, N., Zhang, Y., Wallace, Z. S., et al. (2021). COVID-19 outcomes in patients with systemic autoimmune rheumatic diseases compared to the general population: a US multicenter, comparative cohort study. *Arthritis and Rheumatology* 73 (6), 914–920. doi:10.1002/art.41619
- Ebeling, M., Derka, S., Baudrexel, J., Sakkas, A., Scheurer, M., Wilde, F., et al. (2024). Ibuprofen reduces 5-year overall survival of head and neck cancer patients with immunotherapy – a retrospective case-controlled real-world data analysis of 10,000 patients. *Anticancer Res.* 44 (1), 313–322. doi:10.21873/anticancer.16814
- Elze, M. C., Gregson, J., Baber, U., Williamson, E., Sartori, S., Mehran, R., et al. (2017). Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. *J. Am. Coll. Cardiol.* 69 (3), 345–357. doi:10.1016/j.jacc.2016.10.060
- Gaertner, F., Preissner, S., Heiland, M., Preissner, R., and Wüster, J. (2024). Beneficial effect of metformin on the five-year survival in about 40,000 patients with head and neck cancer. *Cancers* 16 (5), 982. doi:10.3390/cancers16050982
- Gao, Z., Gorenflo, M., Kaelber, D. C., Monnier, V. M., and Xu, R. (2023b). Drug repurposing for reducing the risk of cataract extraction in patients with diabetes mellitus: integration of artificial intelligence-based drug prediction and clinical corroboration. *Front. Pharmacol.* 14, 1181711. doi:10.3389/fphar.2023.1181711
- Gao, Z., Winhusen, T. J., Gorenflo, M., Ghitza, U. E., Davis, P. B., Kaelber, D. C., et al. (2023a). Repurposing ketamine to treat cocaine use disorder: integration of artificial intelligence-based prediction, expert evaluation, clinical corroboration and mechanism of action analyses. *Addiction* 118 (7), 1307–1319. doi:10.1111/add.16168
- Gershater, B., Bieber, K., Vorobyev, A., Ludwig, M. A., Zirpel, H., De Luca, D. A., et al. (2024). Differential risks of psoriatic arthritis development in patients with varied psoriasis manifestations: a sex- and ethnicity-specific analysis. *Front. Med.* 11, 1385491. doi:10.3389/fmed.2024.1385491
- Ghosh, A., Leong, J. Y., Foss, H., Ebbott, D., Prebay, Z., Li, M., et al. (2024). Evaluating treatment pathways for Peyronie's disease utilizing the TriNetX research registry. *Int. J. Impot. Res.* 36 (2), 168–170. doi:10.1038/s41443-023-00815-0
- Grambsch, P. M., and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81 (3), 515–526. doi:10.1093/biomet/81.3.515
- Griffiths, C. E. M., Armstrong, A. W., Gudjonsson, J. E., and Barker, JNWN (2021). Psoriasis. *Lancet* 397 (10281), 1301–1315. doi:10.1016/S0140-6736(20)32549-6
- Hariton, E., and Locascio, J. J. (2018). Randomised controlled trials – the gold standard for effectiveness research: study design: randomised controlled trials. *BJOG* 125 (13), 1716. doi:10.1111/1471-0528.15199
- Hasan Ali, O., Bomze, D., Ring, S. S., Berner, F., Fässler, M., Diem, S., et al. (2020). BP180-specific IgG is associated with skin adverse events, therapy response, and overall survival in non-small cell lung cancer patients treated with checkpoint inhibitors. *J. Am. Acad. Dermatol.* 82 (4), 854–861. doi:10.1016/j.jaad.2019.08.045
- Heinrich Heine Universität Düsseldorf (2020). G\*Power, statistical power analyses for mac and windows. Available at: <https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>.
- Henao-Martinez, A. F., Chastain, D. B., and Thompson, G. R. (2023). “Treatment pathways, switches, and inappropriate treatment during invasive pulmonary aspergillosis: real-world experiences from a global research network study.”. Editor D. J. Krysan, 67, e0072123–23. doi:10.1128/aac.00721-23Antimicrob. Agents Chemother.
- Hernán, M. A., Hernández-Díaz, S., and Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology* 15 (5), 615–625. doi:10.1097/01.ede.0000135174.63482.43
- Hernán, M. A., and Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available: table 1. *Am. J. Epidemiol.* 183 (8), 758–764. doi:10.1093/aje/kwv254
- Jia, K., Kundrot, S., Palchuk, M. B., Warnick, J., Haapala, K., Kaplan, I. D., et al. (2023). A pancreatic cancer risk prediction model (Prism) developed and validated on large-scale US clinical data. *eBioMedicine* 98, 104888. doi:10.1016/j.ebiom.2023.104888
- Kaakour, A. H., Hua, H. U., and Rachitskaya, A. (2022). Representation of Race and ethnicity in randomized clinical trials of diabetic macular edema and retinal vein occlusion compared to 2010 US census data. *JAMA Ophthalmol.* 140 (11), 1096–1102. doi:10.1001/jamaophthalmol.2022.3929
- Kasperkiewicz, M., Elbuluk, N., Ngo, B. T., Kridin, K., and Ludwig, R. J. (2025). Differing features of pemphigus in skin of colour versus white skin: a broad-scale cohort study. *Acad. Dermatol. Venereol.* 39, e54–e56. doi:10.1111/jdv.20094
- Kasperkiewicz, M., Vorobyev, A., Bieber, K., Kridin, K., and Ludwig, R. J. (2023). Risk of comorbid autoimmune diseases in patients with immunobullous disorders: a global large-scale cohort study. *J. Am. Acad. Dermatol.* 89 (6), 1269–1271. doi:10.1016/j.jaad.2023.07.1030
- Kim, J., and Seo, B. S. (2013). How to calculate sample size and why. *Clin. Orthop. Surg.* 5 (3), 235–242. doi:10.4055/cios.2013.5.3.235
- Kovvuru, S., Nalleballe, K., Onteddu, S. R., Sharma, R., Jasti, M., Kapoor, N., et al. (2021). Immunosuppression in chronic autoimmune neurological disorders during the COVID-19 pandemic. *J. Neurological Sci.* 420, 117230. doi:10.1016/j.jns.2020.117230
- Kridin, K., Abdelghaffar, M., Mruwat, N., Ludwig, R. J., and Thaci, D. (2024b). Are interleukin 17 and interleukin 23 inhibitors associated with malignancies? insights from an international population-based study. *Acad. Dermatol. Venereol.* 38 (2), 315–324. doi:10.1111/jdv.19520
- Kridin, K., Bieber, K., Vorobyev, A., Moderegger, E. L., Hernandez, G., Schmidt, E., et al. (2024a). Risk of death, major adverse cardiac events and relapse in patients with bullous pemphigoid treated with systemic or topical corticosteroids. *Br. J. Dermatol.* 191 (4), 539–547. doi:10.1093/bjd/ljae219
- Kridin, K., and Ludwig, R. J. (2023a). Isotretinoin and the risk of psychiatric disturbances: a global study shedding new light on a debatable story. *J. Am. Acad. Dermatol.* 88 (2), 388–394. doi:10.1016/j.jaad.2022.10.031
- Kridin, K., and Ludwig, R. J. (2023b). Isotretinoin and the risk of inflammatory bowel disease and irritable bowel syndrome: a large-scale global study. *J. Am. Acad. Dermatol.* 88 (4), 824–830. doi:10.1016/j.jaad.2022.12.015
- Kridin, K., and Ludwig, R. J. (2024). Isotretinoin is not associated with inflammatory bowel disease even when compared with topical antibiotics and topical retinoids. *J. Am. Acad. Dermatol.* 90 (4), e137–e138. doi:10.1016/j.jaad.2023.11.004

- Kridin, K., Mruwat, N., Thaci, D., and Ludwig, R. J. (2023b). The cardiometabolic safety of interleukin 23 versus interleukin 17 inhibitors in psoriasis: a large-scale global cohort study. *Acad. Dermatol. Venereol.* 37, e683–e685. doi:10.1111/jdv.18830
- Kridin, K., Vorobyev, A., Papara, C., De Luca, D. A., Bieber, K., and Ludwig, R. J. (2023c). Risk factors and sequelae of epidermolysis bullosa acquisita: a propensity-matched global study in 1,344 patients. *Front. Immunol.* 13, 1103533. doi:10.3389/fimmu.2022.1103533
- Kridin, K., Zirpel, H., Mruwat, N., Ludwig, R. J., and Thaci, D. (2023a). Evaluating the risk of infections under interleukin 23 and interleukin 17 inhibitors relative to tumour necrosis factor inhibitors – a population-based study. *Acad. Dermatol. Venereol.* 37 (11), 2319–2326. doi:10.1111/jdv.19328
- Liu, F., and Panagiotakos, D. (2022). Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Med. Res. Methodol.* 22 (1), 287. doi:10.1186/s12874-022-01768-6
- Makady, A., De Boer, A., Hillege, H., Klungel, O., Goetsch, W., and on behalf of GetReal Work Package 1 (2017). What is real-world data? A review of definitions based on literature and stakeholder interviews. *Value Health* 20 (7), 858–865. doi:10.1016/j.jval.2017.03.008
- Matthews, A. A., Danaei, G., Islam, N., and Kurth, T. (2022). Target trial emulation: applying principles of randomised trials to observational studies. *BMJ* 378, e071108. doi:10.1136/bmj-2022-071108
- Moradi Tuchayi, S., Makrantonaki, E., Ganceviciene, R., Dessinioti, C., Feldman, S. R., Zouboulis, C. C., et al. (2015). Acne vulgaris. *Nat. Rev. Dis. Primers.* 1, 15029. doi:10.1038/nrdp.2015.29
- Murphy, L. R., Hill, T. P., Paul, K., Talbott, M., Golovko, G., Shaltoni, H., et al. (2023). Tenecteplase versus alteplase for acute stroke: mortality and bleeding complications. *Ann. Emerg. Med.* 82 (6), 720–728. doi:10.1016/j.annemergmed.2023.03.022
- Olbrich, H., Kridin, K., Hernández, G., Zirpel, H., Sadik, C. D., Terheyden, P., et al. (2024). Increased cardiovascular risks and mortality in prurigo nodularis: a global cohort study. *EBioMedicine* 103, 105123. doi:10.1016/j.ebiom.2024.105123
- Olbrich, H., Kridin, K., Zirpel, H., Sadik, C. D., Terheyden, P., Thaçi, D., et al. (2023). Cutaneous lupus erythematosus is associated with an increased risk of cardiac and vascular diseases: a large-scale, propensity-matched global retrospective cohort study. *EBioMedicine* 93, 104639. doi:10.1016/j.ebiom.2023.104639
- Ozhathil, D. K., Powell, C. M., Corley, C. V., Golovko, G., Song, J., El Ayadi, A., et al. (2024). A retrospective analysis of mortality risk and immunosuppressive therapy for Stevens-Johnson Syndrome and toxic epidermal necrolysis syndrome using the TriNetX research network. *Burns* 50 (1), 75–86. doi:10.1016/j.burns.2023.08.009
- Palchuk, M. B., London, J. W., Perez-Rey, D., Drebert, Z. J., Winer-Jones, J. P., Thompson, C. N., et al. (2023). A global federated real-world data and analytics platform for research. *JAMIA Open* 6 (2), ooad035. doi:10.1093/jamiaopen/ooad035
- Patel, K., Waldron, D., and Graziane, N. (2023). Re-purposing FDA-approved drugs for opioid use disorder. *Subst. Use and Misuse* 58 (13), 1751–1760. doi:10.1080/10826084.2023.2247071
- Raap, U., Limberg, M. M., Kridin, K., and Ludwig, R. J. (2023). Pruritus is associated with an increased risk for the diagnosis of autoimmune skin blistering diseases: a propensity-matched global study. *Biomolecules* 13 (3), 485. doi:10.3390/biom13030485
- Ramos-Casals, M., Brahmer, J. R., Callahan, M. K., Flores-Chávez, A., Keegan, N., Khamashta, M. A., et al. (2020). Immune-related adverse events of checkpoint inhibitors. *Nat. Rev. Dis. Prim.* 6 (1), 38. doi:10.1038/s41572-020-0160-6
- Reynolds, R. V., Yeung, H., Cheng, C. E., Cook-Bolden, F., Desai, S. R., Druby, K. M., et al. (2024). Guidelines of care for the management of acne vulgaris. *J. Am. Acad. Dermatol.* 90 (5), 1006.e1–1006.e30. doi:10.1016/j.jaad.2023.12.017
- Ribas, A., and Wolchok, J. D. (2018). Cancer immunotherapy using checkpoint blockade. *Science* 359 (6382), 1350–1355. doi:10.1126/science.aar4060
- Riley, D. R., Essa, H., Austin, P., Preston, F., Kargbo, I., Ibarburu, G. H., et al. (2023). All-cause mortality and cardiovascular outcomes with sodium-glucose Co-transporter 2 inhibitors, glucagon-like peptide-1 receptor agonists and with combination therapy in people with type 2 diabetes. *Diabetes. Obes. Metab.* 25 (10), 2897–2909. doi:10.1111/dom.15185
- Rivera-Caravaca, J. M., Buckley, B. J. R., Harrison, S. L., Fazio-Eynullayeva, E., Underhill, P., Marin, F., et al. (2021). Direct-acting oral anticoagulants use prior to COVID-19 diagnosis and associations with 30-day clinical outcomes. *Thrombosis Res.* 205, 1–7. doi:10.1016/j.thromres.2021.06.014
- Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology* 1 (1), 43–46. doi:10.1097/00001648-199001000-00010
- Silva, J. D., Jupiter, D. C., and Taghialatela, G. (2024). Reduced prevalence of Parkinson's disease in patients prescribed calcineurin inhibitors. *J. Parkinson's Dis.* 14 (3), 533–543. doi:10.3233/JPD-230313
- Silva, J. D., Taghialatela, G., and Jupiter, D. C. (2023). Reduced prevalence of dementia in patients prescribed tacrolimus, sirolimus, or cyclosporine. *JAD* 95 (2), 585–597. doi:10.3233/JAD-230526
- Singla, S., Putman, M., Liew, J., and Gordon, K. (2023). Association between biological immunotherapy for psoriasis and time to incident inflammatory arthritis: a retrospective cohort study. *Lancet Rheumatology* 5 (4), e200–e207. doi:10.1016/S2665-9913(23)00034-6
- Solovic, I., Sester, M., Gomez-Reino, J. J., Rieder, H. L., Ehlers, S., Milburn, H. J., et al. (2010). The risk of tuberculosis related to tumour necrosis factor antagonist therapies: a TBNET consensus statement. *Eur. Respir. J.* 36 (5), 1185–1206. doi:10.1183/09031936.00028510
- Stappf, M. P. (2018). Using real world data to assess cardiovascular outcomes of two antidiabetic treatment classes. *WJD* 9 (12), 252–257. doi:10.4239/wjd.v9.i12.252
- Sutaria, N., Adawi, W., Brown, I., Parthasarathy, V., Roh, Y. S., Choi, J., et al. (2022). Racial disparities in mortality among patients with prurigo nodularis: a multi-center cohort study. *J. Am. Acad. Dermatol.* 86 (2), 487–490. doi:10.1016/j.jaad.2021.09.028
- Sylvestre, S., Craig, T., Ajewole, O., Craig, S., Kaur, S., and Al-Shaikhly, T. (2021). Racial and ethnic disparities in the research and care of hereditary angioedema patients in the United States. *J. Allergy Clin. Immunol. Pract.* 9 (12), 4441–4449.e2. doi:10.1016/j.jaip.2021.08.019
- Tan, N. K. W., Tang, A., MacAlevey, NCYL, Tan, B. K. J., and Oon, H. H. (2024). Risk of suicide and psychiatric disorders among isotretinoin users: a meta-analysis. *JAMA Dermatol.* 160 (1), 54–62. doi:10.1001/jamadermatol.2023.4579
- Tang, K., Seo, J., Tiu, B. C., Le, T. K., Pahalyants, V., Raval, N. S., et al. (2022). Association of cutaneous immune-related adverse events with increased survival in patients treated with anti-programmed cell death 1 and anti-programmed cell death ligand 1 therapy. *JAMA Dermatol.* 158 (2), 189–193. doi:10.1001/jamadermatol.2021.5476
- Taquet, M., Dercon, Q., Todd, J. A., and Harrison, P. J. (2024). The recombinant shingles vaccine is associated with lower risk of dementia. *Nat. Med.* 30 (10), 2777–2781. doi:10.1038/s41591-024-03201-5
- Taquet, M., Geddes, J. R., Husain, M., Luciano, S., and Harrison, P. J. (2021). 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry* 8 (5), 416–427. doi:10.1016/S2215-0366(21)00084-5
- Taquet, M., Sillett, R., Zhu, L., Mendel, J., Camplisson, I., Dercon, Q., et al. (2022). Neurological and psychiatric risk trajectories after SARS-CoV-2 infection: an analysis of 2-year retrospective cohort studies including 1 284 437 patients. *Lancet Psychiatry* 9 (10), 815–827. doi:10.1016/S2215-0366(22)00260-7
- Taylor, M. T., Margolis, D. J., Kwatra, S. G., and Barbieri, J. S. (2023). A propensity score matched cohort study identifying an association of acne, but not oral antibiotic or isotretinoin use, with risk of incident inflammatory bowel disease. *J. Am. Acad. Dermatol.* 88 (4), 841–847. doi:10.1016/j.jaad.2023.01.014
- Topaloglu, U., and Palchuk, M. B. (2018). Using a federated network of real-world data to optimize clinical trials operations. *JCO Clin. Cancer Inf.* 2 (2), 1–10. doi:10.1200/CCI.17.00067
- TriNetX (2024). How does TriNetX test for proportionality on a hazard ratio. Available at: <https://support.trinetx.com/hc/en-us/articles/360053133594-How-does-TriNetX-test-for-proportionality-on-a-hazard-ratio>.
- TriNetX. Publication Guidelines (2017). TriNetX. Publication guidelines. Available at: <https://trinetx.com/real-world-resources/case-studies-publications/trinetx-publication-guidelines/>.
- U.S. Food and Drug Administration (2024a). Purple book, database of licensed biological products. Available at: <https://purplebooksearch.fda.gov>.
- U.S. Food and Drug Administration (2024b). *Real-world evidence*. Available at: <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>.
- Vell, M. S., Loomba, R., Krishnan, A., Wangenstein, K. J., Trebicka, J., Creasy, K. T., et al. (2023). Association of statin use with risk of liver disease, hepatocellular carcinoma, and liver-related mortality. *JAMA Netw. Open* 6 (6), e2320222. doi:10.1001/jamanetworkopen.2023.20222
- Wang, W., Volkow, N. D., Berger, N. A., Davis, P. B., Kaelber, D. C., and Xu, R. (2024a). Association of semaglutide with risk of suicidal ideation in a real-world cohort. *Nat. Med.* 30 (1), 168–176. doi:10.1038/s41591-023-02672-2
- Wang, W., Volkow, N. D., Berger, N. A., Davis, P. B., Kaelber, D. C., and Xu, R. (2024b). Association of semaglutide with tobacco use disorder in patients with type 2 diabetes: target trial emulation using real-world data. *Ann. Intern. Med.* 177 (8), 1016–1027. doi:10.7326/M23-2718
- Wang, W., Volkow, N. D., Berger, N. A., Davis, P. B., Kaelber, D. C., and Xu, R. (2024c). Association of semaglutide with reduced incidence and relapse of cannabis use disorder in real-world populations: a retrospective cohort study. *Mol. Psychiatry* 29 (8), 2587–2598. doi:10.1038/s41380-024-02498-5
- Wang, W., Wang, C. Y., Wang, S. I., and Wei, J. C. C. (2022). Long-term cardiovascular outcomes in COVID-19 survivors among non-vaccinated population: a retrospective cohort study from the TriNetX US collaborative networks. *eClinicalMedicine* 53, 101619. doi:10.1016/j.eclinm.2022.101619
- Wright, S., Strunk, A., and Garg, A. (2021). Risk of new-onset inflammatory bowel disease among patients with acne vulgaris exposed to isotretinoin. *J. Am. Acad. Dermatol.* 84 (1), 41–45. doi:10.1016/j.jaad.2020.07.042
- Wüster, J., Heiland, M., Nahles, S., Preissner, R., and Preissner, S. (2023). Statin medication improves five-year survival rates in patients with head and neck cancer: a retrospective case-control study of about 100,000 patients. *Cancers* 15 (12), 3093. doi:10.3390/cancers15123093