RWE POSITION PAPER

Real World Evidence (RWE): An Indian perspective

AUTHORS:

Real World Evidence Council working group (2021-22)

AFFILIATIONS:

This position paper was written by the Real World Evidence council working group of Indian Society Clinical Research.

CORRESPONDING AUTHOR DETAILS:

Dr Deepa Chodankar, Heading GENESIS unit, Sanofi

Address: L&T Business Park, Saki Vihar Road, Powai
Mumbai 400 072
email: deepa.chodankar@sanofi.com

RELEASED ON: 10 MAY 2022

ABOUT ISCR:

The Indian Society for Clinical Research (ISCR) is an association of Indian clinical research professionals registered under the Societies Registration Act (1860) in India. The Society brings together all those who are engaged in clinical research activities in India and provides a forum for the exchange of information and learning. ISCR aims to build awareness of clinical research as a specialty in India and to facilitate its growth in the country while helping to evolve the highest standards of quality and ethics.

REAL WORLD EVIDENCE WORKING GROUP:

The ISCR RWE Working Group (2022) is an independent group formed from the RWE council (ISCR) to evaluate the current state, challenges, benefits, opportunities and regulations for the conduct of Real-world evidence studies (RWE) in India. The objective of this working group was to evaluate Indian landscape for RWE studies and to provide guidance on how to conduct a RWE study in India.

CONFIDENTIALITY STATEMENT:

The information presented in this position paper draws upon the combined current understanding and knowledge of ISCR RWE council Working Group and is provided as an aid to evaluate the conduct on RWE studies in India. The opinions of the author(s), ISCR, and the RWE council Working Group do not necessarily reflect the position of individuals or organizations (academia/ industry). Users should assess the contents and opinions in the light of their own knowledge, needs, and experience as well as interpretation of relevant guidance and regulations.

LICENSE FOR USE OF THIS CONTENT:

CITED AS:

Available from: https://www.iscr.org/Articles.asp

Table of Contents

1 EXECUTIVE SUMMARY

The healthcare sector is rapidly evolving in response to the exponential growth in the volume and diversity of patient information. With rising interest of healthcare industry on patient outcomes, the information beyond randomized clinical trials (RCTs) has become significant and is anticipated to add more value to existing clinical evidence. Real-world data (RWD) is a source of such information and refers to all the information pertaining to an individual's health status and the delivery of health, that is collected from varied sources.

The objective of this paper is to evaluate Indian landscape for RWE studies and to provide guidance on how to conduct RWE study in India. The paper explores various aspects of RWE studies such as RWE study designs, conduct of RWE studies, informed consent process, regulatory requirements in India, analysis and statistical considerations of RWE studies, quality, data privacy and security and reporting of these studies.

2 INTRODUCTION

The healthcare sector is rapidly evolving in response to the exponential growth in the volume and diversity of patient information. The process of clinical decision and patient health outcome are strongly dependent upon factors such as genomics, behavioral, social, and environmental factors. With rising interest of healthcare industry on patient outcomes, the information beyond randomized clinical trials (RCTs) has become significant and is anticipated to add more value to existing clinical evidence. Real-world data (RWD) is a source of such information and refers to all the information pertaining to an individual's health status and the delivery of health, that is collected from varied sources [1], including but not limited to electronic health records (EHRs), insurance data, product and disease registries. Advantages derived from RWD include the availability of timely data at reasonable cost, large sample sizes that enable analysis of subpopulations and less common effects, and the representativeness of real-world practice and behaviors. The analysis and usage of RWD to generate valuable insights and a better understanding of potential risks and benefits of a medical product is called as real-world evidence (RWE) [2].

An RWE study systematically blends the clinical, regulatory, and commercial aspects of clinical research and public health domain [3]. The RWE has the potential to complement the information gathered from conventional clinical trials that are usually based on specific research environment. It provides information on drug development, disease history and progression, patient's response to treatment, effectiveness and outcomes research, research on health care infrastructure, quality improvement, and safety surveillance. It also presents impact of other exogenous factors such as clinical setting and health-system characteristics on therapeutic efficacy and patient outcomes.

Often RCTs fail to provide a comprehensive assessment of a medical product as they are based on relatively small patient population in controlled environments [4]. The U.S. Food and Drug Administration (FDA) acknowledged a larger role for the use of RWD and RWE to support regulatory affairs [1]. An RWE study mandates review by the ethics committee, and requires informed consent of potential participants to access medical records [2]. The significance of RWE lies in its potential to provide a more comprehensive understanding of how a new medical strategy will function in the "real world" in addition to traditional RCTs.

For instance, during the COVID-19 pandemic, the actual potential of RWD was harnessed and rapid, actionable insights were available to understand, such as: disease etiology, efficacy and safety of treatments, health-care decisions, and policies [5, 6]. A study on COVID-19 deaths (n=5683) in Great Britain demonstrated a strong association between mortality and gender, age and nutritional status, ethnicity, uncontrolled diabetes, severe asthma, obesity, chronic heart disease, liver disease, stroke/ dementia, other neurological diseases, reduced kidney function, autoimmune diseases, malignancy, and hypertension [7]. In another retrospective study, a cohort of hospitalized COVID-19 patients was analyzed to develop a predictor risk score based on factors

such as chest X-ray assessment, age, hemoptysis, dyspnea, unconsciousness, evidence of comorbidities, history of cancer, neutrophil-to-lymphocyte ratio, lactate dehydrogenase, and direct bilirubin [8]. These RWE studies have demonstrated promising clinical implications for the management of COVID-19 and facilitated critical inquiry on efficacy and safety profile of treatments and health economics and outcomes research.

Besides clinical impact, the regulators can utilize RWE as a tool to monitor post-marketing safety assessments, and aid FDA in decision-making [9]. This will predominantly impact drug approvals and accelerate the tedious review process during drug development. For healthcare providers and clinicians, RWE studies may assist in deriving clinical guidelines [10]. By using RWE, we can enhance our understanding of what works for different patients. Real world evidence allows researchers to examine the performance of drug treatments and other interventions while also looking at other factors and variables. The findings can build and evolve the understanding of a disease and help in treatment decisions. Thus, RWE can aid physicians to design and execute more individualized and adequate therapeutic strategy for management of patients.

The healthcare providers form an integral component of RWD collection system as they routinely compute EHR including structured clinical data such as diagnoses, diagnostic assessments, and prescribed drugs. Real world data also include medical claims that healthcare providers forward to the insurance firms.

In addition, RWE studies are relatively fast and cost effective compared to standard RCTs. The patient sub-groups not included in the specifically designed RCTs can be effectively analysed in RWE studies for more rigor and can present novel insights to the available evidence. For example, with RWE, researchers can study how new therapies work among patients with co-morbidities, certain age groups, or specific socio-demographic groups. Real world evidence can analyse patient information over a lifetime and not just during a specified period.

Real world evidence studies can utilize artificial intelligence and machine learning based patient-monitoring systems [11]. These tools employ images and videos from wearable sensory devices and record and process data while attached to the human body, without any need of a hardwired connection [12]. Machine learning models, coupled with wearable devices, have been applied in automatic detection of cognitive and emotional states, in monitoring participants in Parkinson's disease trials, and in assessing quality of sleep in neurology trials [13]. Machine learning, natural language processing, and optical character recognition could help in analyzing unstructured medical records for RWE studies [11].

Although methods and algorithms for RWD analysis are being designed and optimized, the utility of RWE is marred by technical snags such as confounding factors, quality and heterogeneity of data, and bias [14]. For India, there are additional inevitable challenges such as pattern of Indian clinical practice and interest of physicians in RWE studies [2]. Adequate patient management and follow ups are inconsistent and depends upon patients' socio-economic and educational background and on the awareness about his/ her medical condition. This leads to erratic

assessment of treatment efficacy and safety. This is further complicated by insufficient documentation and medical history records [15]. Moreover, healthcare professionals often do not show enough motivation or commitment during recruitment in structured RWE registries.

Thus, it is of utmost importance to distinguish two key dimensions of RWE. The first dimension is the environment or setting in which RWE is generated, i.e., the population providing the data and the specific methods employed for data collection, mining, and curation; the second dimension is the approach used to conduct the surveillance or research [3].

3 REAL-WORLD EVIDENCE STUDY DESIGNS

The study designs for RWE studies can be described according to different perspectives:

- The framework as defined formally or informally by regulatory bodies, scientific societies or through initiatives that have been created to promote the understanding and application of RWE in health science
- The type of data use
- The time frame of the study
- The assignment of exposure

There is no uniform framework to classify the RWE study designs, hence the different perspectives are briefly presented in this section.

3.1 Regulatory bodies, scientific societies, health technology agencies definitions

1. **US FDA:** The 21st Century Cures Act, passed in 2016, placed additional focus on the use of these types of data to support regulatory decision making, including approval of new indications for approved drugs. Congress defined RWE as data regarding the usage, or the potential benefits or risks, of a drug derived from sources other than traditional clinical trials. The US FDA has expanded on this definition as: RWE is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD. Real world evidence can be generated by different study designs or analyses, including but not limited to [9, 33]:

    - Randomized trials (e.g., large simple trials, pragmatic clinical trials) and
    - Observational studies (prospective or retrospective)

    In the framework for RWE program intended for application to biological products licensed under the Public Health Service Act, FDA further defines:

    - A clinical trial as a research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes.
    - Non-randomized, Single Arm Trials with External RWD Control. Typically, the external control arm uses data from past traditional clinical trials, but in some cases, RWD have been used as the basis for external controls.
    - Observational studies are non-interventional clinical study designs that are not considered clinical trials.

2. **EMA:** The European Medicines Agency (EMA) doesn't have a framework on RWE like the US FDA. In 2018, EMA published a Regulatory Perspective on RWE in scientific advice [16]. In this document, RWE designs mentioned include Post Approval Efficacy Studies (PAES), Post Approval Safety Studies (PASS), primary research data collected on how interventions are used in routine clinical practice, secondary research data derived from routinely collected data for other purposes, including pragmatic randomised controlled trials and registries. In September 2020, the EMA released a draft guideline on registry-based studies to provide manufacturers with "recommendations on key methodological aspects that are specific to the use of patient registries" [17].
3. **Health Canada** published in 2019 "Optimizing the Use of Real World Evidence to Inform Decision Making" [18]. Health Canada focuses on the Real World Data/ Evidence Quality aspects rather than the types of study designs
4. **The GetReal Institute** is a European initiative that brings together a wide variety of stakeholders to drive the sustainable development and adoption of tools, methods and best practices in the generation and use of RWE for better health care decision-making. The Get Real Institute developed The RWE Navigator, which has been designed for a wide variety of users. Patients and patient organisations may use it to better understand RWE concepts and the challenges involved in using or generating RWE. Regulators or health technology agencies (HTA) professionals may use it to understand more about the need for RWE and the challenges of designing studies to meet this need [19].
5. **Academic/ HTA bodies frameworks**: several societies (ISPE, ISPOR) and HTA bodies (NICE) also developed frameworks for RWE designs/ methodology classification [20, 21, 22, 23, 24]. Generally, the 3 main types of design are:
    - Primary data collection
    - Secondary data use of data collected in routine practice
    - Hybrid designs

- There is no Indian framework that defines the different designs for RWE studies

3.2 Study Designs according to the source of data (Primary or Secondary):

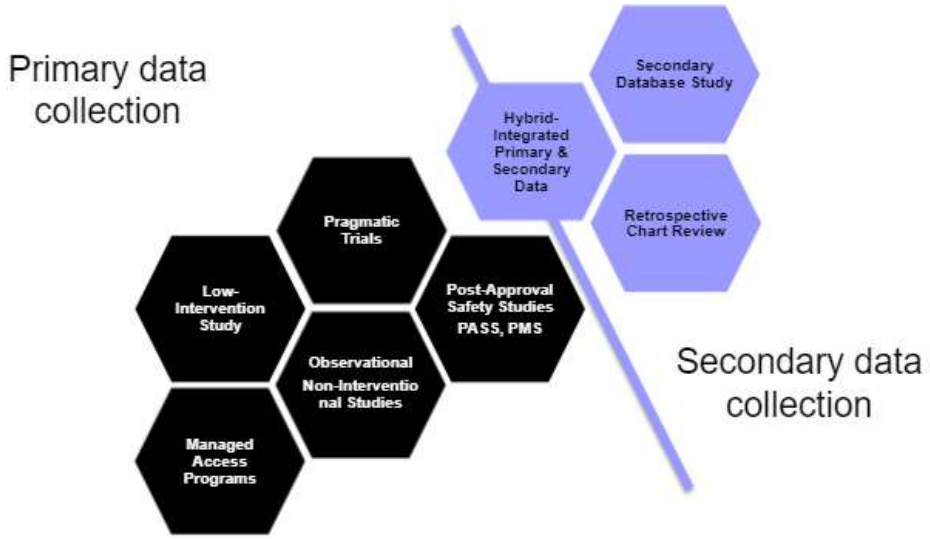

Figure 1

3.2.1 Primary data collection

Primary data are newly generated by an investigator during a study

- Observational study: studies where researchers observe the effect of a risk factor, diagnostic test, treatment or other intervention without trying to change who is or isn't exposed to it. Cohort studies, case control studies and cross-sectional studies are 3 types of observational studies.
- Pragmatic trial/ low intervention trial: a study where patients are randomised to receive an intervention (A vs B) but where the observation of the effect of the intervention is evaluated in a study setting that is similar to what would be experienced in real life.
- PASS (post-approval safety study): a study that is carried out after a medicine has been authorised to obtain further information on a medicine's safety, or to measure the effectiveness of risk-management measures. This may be a regulatory requirement (e.g.: EMA) [25].
- PMS (post-marketing surveillance) study: an open study where unlike pre-marketing studies, the selection of patients is not strictly defined by stringent inclusion and exclusion criteria, but governed by the permissible indications and contra-indications of the drug as stated in the text of prescribing information [26]. This ensures that information is collected in a varied spectrum of patients, and makes it likely that the study will yield data that may not have been captured in Phase III studies. This may be a regulatory requirement (ex: Japan).

3.2.2 Secondary data collection

Secondary data are collected for other purposes and are being used again to answer another research question or hypothesis

- Secondary database study: Secondary data are generated for their primary purpose (e.g.: claims database, disease registry, Electronic Medical Records (EMR), etc. The consequence of that provenance is that they follow their own languages, structure, platforms and data formats. Access or dissemination, even for legitimate research purposes, is tightly limited by their primary owners and governed by rules and regulations for personal data protection. It is also not guaranteed that any RWD from such sources are fit for a specific research purpose.
- Retrospective chart reviews: The retrospective chart review (RCR), also known as a medical record review, is a type of research design in which pre-recorded, patient-centered data are used to answer one or more research questions [27]. The data used in such reviews exist in many forms: electronic databases, results from diagnostic tests, and notes from health service providers to mention a few.

3.2.3 Hybrid integrated Primary and Secondary Data

Hybrid or enriched RWD studies combine primary data collected directly from physicians and patients with existing (secondary) data such as EMRs, insurance claims or established registries. This approach takes advantage of the strengths of both primary and secondary sources, yielding maximum scientific benefit while allowing sponsors to answer more questions within a single study [28].

3.3 Time Frame (relative to the study start or index date)

As opposed to interventional clinical trials which are by nature prospective, RWE studies do not require the allocation of the intervention that is studied (one exception is the Pragmatic Trial

design, see here), thus they can be designed and adopt 3 different time frames to answer a research question: prospective, retrospective or cross-sectional [29].

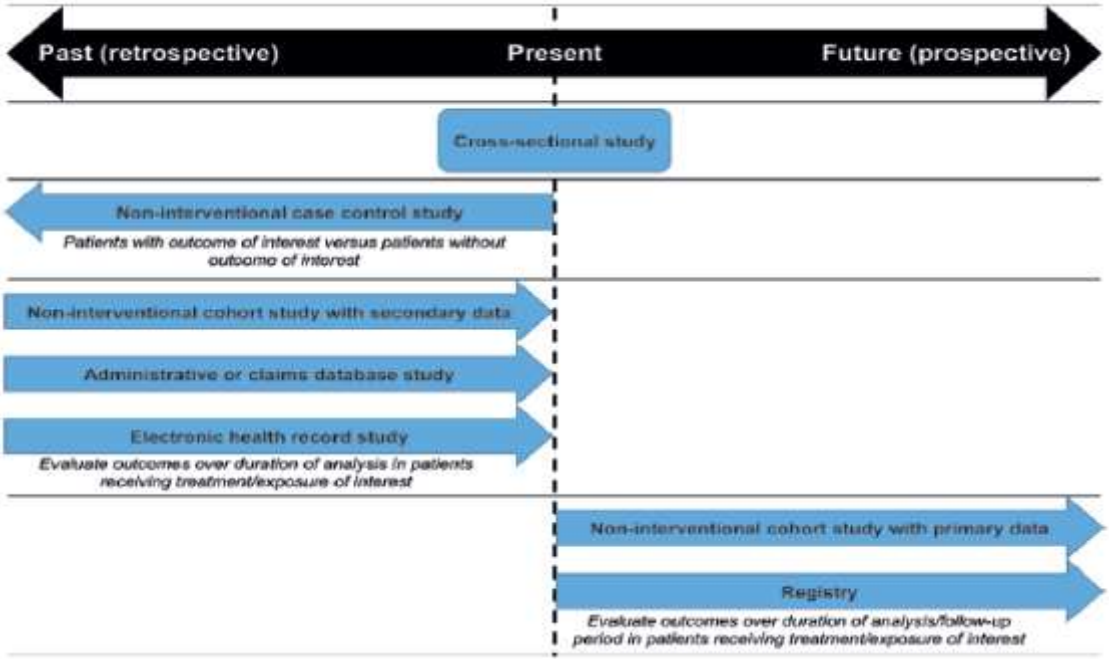

Figure 2

1. In prospective cohort studies/ registries, patients are monitored until a relevant outcome occurs (eg, a major bleeding event, stroke), with data routinely collected on potential risk factors for the outcome.
2. In retrospective cohort studies, the methodology is the same but the data have already been collected for a separate purpose and a post-hoc analysis is carried out. Retrospective cohort studies can be conducted quickly and inexpensively compared with RCTs because the data have already been collected. However, retrospective cohort studies can suffer from limited or missing data, less rigour in data collection and recall bias.

The advantages of longitudinal (prospective or retrospective) cohort studies include the ability to assess a broad range of risk factors, and the fact that several outcomes can be monitored simultaneously. The chronology of the study also enables a clear distinction between cause and effect (unlike cross-sectional studies), although this also means that loss to follow-up can significantly affect the outcomes, and studying rare outcomes can be inefficient.

3. Cross-sectional studies involve the assessment of a single group of patients at a single point in time, at which treatment and outcomes are determined simultaneously. They are typically used to assess prevalence and infer the cause of conditions/ outcomes. Cross-sectional studies can be conducted relatively quickly and inexpensively compared with RCTs, and can assess multiple outcomes simultaneously. They are, therefore, the

most efficient way to determine the prevalence of a condition. However, because the data are collected at a single time point, it is difficult to clearly distinguish cause and effect, e.g., patients who develop an outcome but die before the end of the study are not captured, and they are susceptible to selection bias.

4. Case-control studies: are usually conducted retrospectively. Patients who have experienced the outcome of interest are matched with a control group who have not experienced this outcome, and exposure to treatment or other factors are assessed from medical history to determine causality. Because the patient population is selected based on the outcome, case-control studies are especially useful in studying rare conditions or those with a long latency between exposure and disease. They can also consider many variables simultaneously, providing a case-efficient way of identifying potential predictors of specific outcomes. However, case-control studies are susceptible to sampling bias, observational bias and recall bias, as well as unmeasured confounders.

3.4 Exposure/ intervention assignment

Another perspective is if the exposure/ intervention under study is assigned by the investigator or not. In most RWE designs (case-control, cohort, cross-sectional, analysis of claims or EMR, registries) it is not, the only exception is the so-called Pragmatic Trial [30].

CLASSIFICATION OF RWE STUDY DESIGN

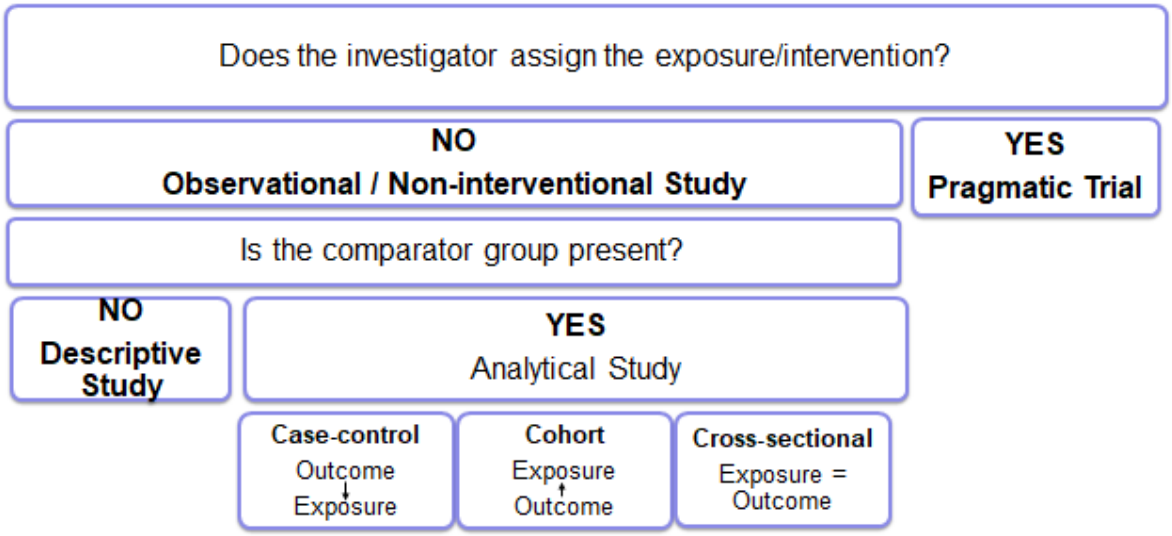

Figure 3

Pragmatic trials aim to measure the relative effectiveness of treatment strategies in real-world clinical practice. These are also called Low Interventional Studies. They provide evidence of the added value of a treatment strategy in routine clinical practice, while maintaining the strength of comparisons based on randomised controlled trials. In a pragmatic trial a comparison is made

between randomised groups of patients that are similar to the target group in the characteristics that influence outcome event rates and potentially modify drug response, in a study setting that is similar to what would be experienced in real life. The treatment strategies being compared, and the outcome measures used for the comparison should be relevant to routine clinical practice. The term 'pragmatic trial' is commonly used for trials that assess the difference between treatment strategies, to include the potential impact of extraneous factors other than the pharmacological effect of the medicine (such as co-medication, non-adherence and placebo effects). The aim is to maximise generalisability of the results to a broader setting or patient population, for example the 'decision-making' population being considered by HTA for a reimbursement agency.

3.5 Mixed designs - External Control Arm

Mixed designs RWE combine the strengths of clinical trials and of RWE studies. The "external control arm" is a design that provides a second, 'external' control arm to a single-arm clinical trial [31]. This design is most-often done in rare and life-threatening diseases, including oncology, where it is difficult, or even unethical to enroll a fully separate placebo or standard-of-care arm, or where effective treatment isn't otherwise available. It requires identification of multiples of the treatment arm, to allow for cohort balancing/ matching against the trial population. Careful selection of the matching criteria and covariates is key, it must be clinically relevant. External controls established based on real-world data in natural disease cohorts, particularly in rare diseases) can be historical external controls (based on real-world data obtained earlier) or parallel external controls (based on data from disease registries constructed simultaneously with the single-arm trial).

4 CONDUCT OF REAL-WORLD EVIDENCE STUDIES

Real world studies are designed to collect data in the regular clinical practice, and/ or patient reported outcomes (PRO) to generate evidence for different stake holders from regulators, payers, providers, patients, and the biopharma industry [32, 33].

Key operational considerations to conduct RWE studies in comparison with RCT are as follows

Parameter	RCT	RWE	Operational considerations
✓ Study Sites	✓ Experienced in Clinical Trials with needed infrastructure for study ✓ Limited number ✓ Site Selection Visit On-Site	✓ Large number of sites to ensure appropriate representativeness of actual population ✓ For Secondary data analysis studies, perform date driven site identification for eligible data. ✓ Site Selection Visit On-Site or Remote	✓ Interest to be part of RWE is generally low due to low study grants, no resources, lack of understanding on scientific importance of study- resulting in high % of inexperienced sites. [34] ✓ Ensuring investigators understand the requirements upfront e.g., Study objectives, Informed Consent/ Data Collection is key to successful participation ✓ Inexperienced sites need more support/ guidance on study activities (EC submissions, grant process, Data collection etc.) [35]. ✓ A careful consideration should be taken in the site selection strategy based on the study objective to include over all representativeness of the target population with innovative patient recruitment methodologies, types of data sources be it primary or secondary data type. [36].
✓ Regulation/ Guidance/ Laws ✓ / Ethics	✓ Clear regulatory framework ✓ Clear ethical framework	✓ ICMR Ethics guidelines 2017 does not specifically define RWE studies, it does have a section mentioned “special issues related to datasets” [64].	✓ Will typically require ethical/ IRB approval ✓ In addition, must consider local marketing laws/ guidance as well as data privacy laws.
✓ Study Drug	✓ Investigational or marketed product provided and labeled ‘for	✓ There may be no product, e.g., disease registry, Observational study	✓ Understanding of product use, Market uptake, Off label use is critical for right site selection and ultimate patient recruitment ✓ Sites need to be product users – no provision of product unless specially

Parameter	RCT	RWE	Operational considerations
	clinical trial use only'	✓ Marketed product is prescribed as part of standard clinical practice and labeled as per the marketing authorization	requested by regulators in post-marketing authorization safety studies (PASS). ✓ For the secondary data analysis, there is no study drug involved, i:e data collected for other purposes during routine medical care through electronic medical records (EMR), administrative claims data, registries included.
✓ Patient Enrollment	✓ Often include strict inclusion and exclusion criteria that curtails the patient eligibility to participate into the study.	✓ Studies aim for diverse patients that are treated under routine clinical practices and the purpose is to study the safety and effectiveness of the treatment as per the standard of care. ✓ For secondary data studies, protocol and data planning is designed on the target data collection method, data extraction and integration.	✓ Recruit patients with the relevant indication and treatment as per the protocol following the routine clinical practice. Refer to Physician data bases and or Electronic medical records (EMR) databases to select eligible patients and contact patients through mail outs, advertisements, and other appropriate modes of communication to share the study information. ✓ Real world studies to include all types of patients as per the treatment conditions indicated in protocol. The eligibility criteria should be broad to comprise patient population that are representative of heterogenous groups, vulnerable or special population, different age categories, comorbidities to maximize the generalizability of the data. Patients of contraindications should be excluded however patient should not be selected based on treatment compliance, number of site visits or any other practical constraints. [36] [37, 38] ✓ Site should be carefully trained on the study-specific requirements to assist in the data collection and extraction methods.
✓ Patient retention	✓ Patient retention is crucial, generally patient engagement and retention strategies are well developed	✓ Patient retention is an important aspect for the quality and completeness of the data. ✓ Patient retention may vary	✓ Retention planning should be developed from early stages of study and add to the enrolment risk planning with the necessary steps to reduce the dropout rates and encourage patient completion.

Parameter	RCT	RWE	Operational considerations
	through the duration of the study.	depending on the study type, indication, design, duration, study procedures, patient demographics [39].	✓ Protocol procedures to focus on real world data collection and limit the study participation burden for investigators and patients. ✓ Patient centric study design to enable maximum adherence to protocol. ✓ Fair market value compensation for the time spent on any study visits and procedures. ✓ Patient visit tracking can be helpful to understand routine care visit but refrain from additional reminders as this interfere the understanding of the real-world standard of care. ✓ Arrange patient information sheet that the investigators can use to share information about the study progress with patients to encourage retention, PRO completion, etc. [39]
✓ Data Collection	✓ Prospective ✓ CRF size often large	✓ Prospective/ Retrospective/ combination ✓ CRF size ideally smaller	✓ Data collection needs to fit in with daily routine care and EDC needs to work for research naïve sites ✓ CRF design needs to include options for missing data values (e.g., UNK/ UNK/ YYYY). Analysis plan built to support missing data [39, 40] ✓ Built-in edit checks should be maximized to allow quality control at point of data entry to minimize manual query resolution workload at site [41, 42] ✓ Retrospective studies: Data not originally recorded for research purposes and, therefore, may be lacking in quality and quantity; missing values higher. Statistical plan should include methods to handle missing data without interfering the study objectives. ✓ Secondary data studies include data source evaluation, data extraction, variable mapping, data privacy assessment.
✓ Monitoring Strategy	✓ On-site ✓ Routine and periodic	✓ Targeted, fit for purpose monitoring approach	✓ Simplified approaches and user-friendly technology reduce burden on site so as to not interfere with physicians' routine daily practice.

Parameter	RCT	RWE	Operational considerations
	✓ High frequency as Source Data Verification (SDV) % is generally high		✓ Centralized monitoring, use of call centers/ remote monitoring of sites reduces site management costs and still provides the necessary guidance to physicians. ✓ Centralized monitoring is a remote evaluation of accumulating data, performed in a timely manner, supported by appropriately qualified and trained persons (e.g., clinical operations, data managers, biostatisticians, epidemiologist, medical monitors) [43, 44]. ✓ Real world studies should focus on the minimum required level of SDV, while focusing on holistic data review from protocol writing through the various cross functional quality measures with an intent of proactive and early detection of eligibility, quality, safety and operational risks based on the regular data monitoring and risk assessment [45].

Table 1

5 INFORMED CONSENT

The Nuremberg Code (1949) established for the first time the need for a voluntary consent before an individual participates in a clinical trial [46, 47]. Enshrined in the Declaration of Helsinki, the obligation to obtain informed consent protects the individual's freedom of choice and respects the individual's autonomy [48, 49]. Many guidelines and regulations for clinical trials around the world have incorporated the duty to seek written and prospective informed consent from research participants [44, 50]. To ensure valid and voluntary informed consent for clinical trials involving medicinal products, these trial regulations require that potential participants are informed about the purpose of the study, the fact that it constitutes scientific research, the potential risks and benefits, the trial's procedures, and that participants can withdraw at any time during the study without consequences. ln addition, researchers must ensure that the potential participant has understood the information and decided on participation without having been subjected to coercion or undue influence. Thus, informed consent, protects research participants and secure people's trust in clinical research overall.

Recently approved amendments to the European Directive state that only one type of trial may be exempt from the consent requirements: cluster RCTs in which groups of subjects rather than individual subjects are allocated to receive different approved medicinal products may make use of simplified means to obtain informed consent [51]. Necessary conditions are that there are no interventions other than the standard treatments and that the protocol justifies the reasons for obtaining consent by simplified means. The trial should also classify as a low-intervention trial indicating use of approved products in accordance with the marketing authorization and with minimal risk or burden from additional diagnostic or monitoring procedures. The 2012 Ottawa Statement presents similar conditions for a waiver or alteration of informed consent for cluster RCTs, knowingly, that the research is not feasible without modified consent, and that the research-related procedures do not pose more than minimal risk [52].

Real world evidence is evidence that is generated from health data coming from varied sources besides clinical trials e. g. National registries, health care registries. Electronic medical records, data from hospitals and general physicians, longitudinal cohort studies and biobanks, non-medical sources, such as social media applications, personal health monitoring technologies (e.g., home sensors, wearables) RWE can be generated by different study designs or analyses, including but not limited to, randomized trials, including large simple trials, pragmatic trials, observational studies (prospective and/ or retrospective) and Web-based or App-based online survey using electronic-PRO.

Due to the varied sources of RWD and complex nature there are ethical challenges around ownership, transparency, data provenance, privacy and data discrimination, etc. following should be considered while determining how to preserve the overall objective of consent:

- Individuals need to be informed about the various ways their health data may be collected, stored, curated, shared and used in order to foster transparency and trust.
- Governance policies, technical processes, need to be in place to protect participant's privacy and avoid discrimination.

Different data sources may use different types of consent:

1. **Broad Consent** maybe used to allow a data access committee to decide which researchers can use which data for which research questions e. g. Biobanking or longitudinal studies. This type of consent raises issue of governance, how should decisions be made about who gets to use data and for what purposes. This is resolved by:
    - transparent, accountable and streamlined procedures and
    - engagement methods that help to build and maintain trust and ensure study's activities are aligned with participant's interests.
2. **Implied Consent** maybe used for members of the healthcare team, to use data from subject's routine care. The issue in such type of consent is "reuse" is it permissible for other researchers to reuse the routine data for a different research question. This is solved by:
    - having a dynamic consent for new studies and
    - anonymisation and privacy protecting analysis techniques undertaken for existing studies. [53]

In the US, the recent changes to the Common Rule have made it easier to reuse data collected as part of routine healthcare operations for research purposes including EHR data, with additional categories of studies now exempt from an Institutional Review Board (IRB) oversight. However, to publish research involving data from human subjects, virtually all peer-reviewed journals require that the study must have been reviewed and approved by an IRB or ethics board.

Electronic consent has gained a strong foot hold with the advent of the pandemic wherein teleconsultation with family physicians has become a norm. The electronic consent form is embedded in the tablet/ mobile device, a video of the trial specifics is viewed by the potential subject and later questions are answered to confirm the subject's comprehension of the study related research component. Once the potential subject answers the questions, consent is documented electronically, and counter signed by the research investigator or designee. FDA has also issued the guidance on the use of electronic Informed consent. [54]

In India the New Drugs and Clinical Trial Rules 2019 have not described any special provisions for subject consenting in RWD and RWE or electronic consent thus one would go by the understanding that for any prospective observational study conventional consent must be undertaken while for a retrospective observational study using hospital records, registries etc.

one would rely on the institutional policy and procedures on reuse of the health data for research purposes.[55]

The traditional informed consent model for RCTs has been argued to pose substantial hurdles like (recruitment difficulties, reduced generalizability of the results, and selection bias) to the practicability of pragmatic trials. Four alternative informed consent models: integrated consent, targeted consent, broadcast consent, and a waiver of consent have been proposed.

These alternative consent models each aim at overcoming operational and methodological challenges, while still providing patients all the relevant information they need to make informed decisions. Each consent model, however, relies on different attitudes toward the principle of respect for persons and the related duty to inform patients as well as represents different views on whether the common good, demands moral duties from patients to engage in clinical research.

Deviations from traditional consent have ethical implications that need to be balanced. The relative impracticability needs to be weighed against these ethical implications. To adequately perform such an evaluation, it is essential to expose the reasons why traditional consent would affect a particular pragmatic trial's practicability and examine to what degree proposed alternatives affect patients' rights and could be actual solutions. Further work is needed to establish how a pragmatic trial's impracticability have to be balanced against the research risks, along with other, more normative aspects such as patients' rights, their responsibilities, and duties. [56, 57, 58, 59, 60, 61, 62].

6 REGULATORY REQUIREMENTS - REAL-WORLD EVIDENCE STUDIES – INDIA

The New Drugs and Clinical Trials Rules, 2019, and the Drugs and Cosmetics Act do not define RWD or RWE studies [55, 63]

However, the FDA has provided a simplistic definition for regulatory purposes “Real-world evidence (RWE) is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from the analysis of RWD (Real world Data)” [1]

An RWE can be generated by different study designs or analyses, including but not limited to, randomized trials, including large simple trials, pragmatic trials, and observational studies (prospective and/ or retrospective) [9].

As RWE studies are not specifically defined, precise regulatory requirements too are unclear. However, depending on the type/ design and data source of the RWE study, the appropriate regulatory pathway/ approvals in India can be inferred.

For a new drug (within 4 years of approval), a section has been dedicated in the New drugs and Clinical Trials Rules “Post marketing surveillance study or observational or non-interventional study for active surveillance”, which mandates approval of the protocol by central licensing authority. However, as the study drugs are the part of standard of care/ routine treatment at the discretion of the study investigator, the regulatory provisions and guidelines applicable for clinical trial of a new drug are not applicable in such cases [55]

The Clinical Trials rules however have not defined retrospective Vs prospective, primary Vs secondary data generation, Interventional Vs non-Interventional RWE studies. There is no clarity/ consensus on the regulatory process to be followed for such RWE studies in India. Hence there is need to have specific regulatory definition along with guidance and recommendations for conducting RWE studies in India

6.1 Ethical requirements – real world evidence studies

Although the Indian Council for Medical Research (ICMR) Ethics guidelines 2017 do not specifically define RWE studies, it does have a section mentioned “special issues related to

datasets" [64]. This section has highlighted the ethical considerations while harnessing EMR and similar digital patient records for research.

The key pointers mentioned in this section includes: [64]

1) EMR based studies must follow the expected requirements of any other health-related research with due diligence, including review by an Ethics Committee (EC)
2) EC approval is required to establish legitimacy of the purpose for data mining, access control and about the usefulness of information for particular groups (such as rare disease group)
3) Data privacy, data accuracy, data security, and possibility of legal liability should be ensured when the data is outsourced or sold
4) Auditing could be done to detect misuse

The guidelines also has expressed concerns about Digital health records ranging from risks to individual rights, such as privacy and concerns about autonomy to individuals. Ethical issues related to data security, sharing, rights, benefit sharing and others surrounding big data need to be closely examined.

Overall, the guidelines recommends that all reasonable measures must be adopted to respect and protect privacy and confidentiality of individuals

7 ANALYSIS: STATISTICAL CONSIDERATIONS FOR REAL-WORLD EVIDENCE STUDIES

Regulatory agencies, public-private partnerships, and professional organizations have initiated major programs and released guidance or guidelines to address challenges in the use of RWE to inform regulatory decision making [65]. Rigorous and practical methods and practices are needed to define how collecting, analyzing, and reporting RWD should be done. To be influential and useful, RWD needs to be susceptible to robust analytics to confirm that data methods have eliminated biases, controlled quality, and allowed for integration of disparate data sources for both prospective and retrospective studies.

An RWE can be generated by different study designs or analyses, including pragmatic trials, large simple trials, and observational studies (prospective and/ or retrospective). The key data sources include

- **Experimental:** Hybrid or pragmatic clinical trials
- **Non-experimental:** (table below)

Source	Research Data Sources	Transactional data Sources
Purpose	Data Collected primarily for research	Secondary research data
Example	Data specifically for study purpose • Framingham Heart Study • Cardiovascular Health Study • Data intended for other studies • Nurses' Health Study • Some registries	• Clinical documentation • Electronic health records • Wearable devices • Administrative • Claims data • Geocoding/ census

Table 2

Key estimates including population, endpoints, intercurrent events and reporting measures are different in real world studies as compared to RCTs.

Attributes	RCTs	Real World Studies
Population	Detailed inclusion/ exclusion criteria; more homogeneous population & care following protocol	Broad & heterogeneous population from routine clinical practice but: • Confounded by local reimbursement (treatment decision is typically ahead of study participation) and data sources • Require methods to control confounding
Endpoints	Well-defined outcomes measured for the study	• Under-reporting or lack of disease-specific clinical outcomes • Outcome definition/ algorithm with suboptimal specificity & sensitivity • Information bias can be high; uncommon use of natural language processing (NLP) for unstructured data
Intercurrent Events	Extensive efforts devoted for ensure	May be more suitable for studies with long-term follow up, but:

Attributes	RCTs	Real World Studies
	patients follow-up & data completeness	• Treatment change common & reasoning typically not well-recorded • Treatment adherence lower as medication is not provided • Proportion of missing data & loss to follow up higher
Reporting measures	Mostly population- level group comparison for effect size	Very diverse measurements depending on research questions (e.g., prevalence/ incidence, disease progression, treatment pattern, disease burden, population-level comparative effectiveness, patient-level treatment response.

Table 3

7.1 Statistical Analysis Plan

Transparency and validity of real-world studies increases by developing statistical analysis plans (SAPs) before data have been accessed to initiate analyses from pre-planned analyses. The real-world studies should have a detailed statistical analysis plan providing details of the statistical methodologies to be used for the analysis of study data. The structure of the SAP of observational studies is similar to that of RCTs and details of each section can be accessed from publication by Hiemstra et al, 2019 [66]. There are only few adjustments needed for development of a SAP for observational studies when compared to a SAP for RCTs and the SAP for RCTs can be used as a base structure for SAPs of observational studies

7.2 Confounding in Real world studies

Retrospective and prospective real-world studies are subject to bias and confounding factors. Several methods have also been developed to reduce the effects of confounding in observational studies, including propensity score matching (PSM) [67]. This method aims to make it possible to compare outcomes of two treatment or management options in similar patients. It does this by reducing the effects of multiple covariates to a single score, the propensity score. Comparison of outcomes across treatment groups of pairs or pools of propensity-score-matched patients can reduce issues such as selection bias. Although a powerful and widely used tool, there are limits to the degree in which propensity score adjustments can control for bias and confounding variables.

For long-term prospective studies or retrospective data collection, propensity scoring is effective at reducing bias. Propensity scoring considers classification of the relationship between treatment assignment and baseline characteristics. Factors, which are different between two treatment groups and are associated with treatment preference, are weighted to estimate the probability of any participant in the cohort being assigned a specific treatment. This estimated propensity score is used to match the participants across two treatment groups. For analysis of outcomes, “matchable” participants are compared, and unmatchable participants are excluded.

7.3 Sample size estimation for Real world studies

The sample sizes of most observational studies are influenced by factors like resources, time restrictions, and convenience and is different from RCTs that calculate a sample size to study an intervention effect taking power into consideration. Accordingly, most observational studies will have a given sample size and, if sufficiently large, affording enough power. The STROBE guidelines only expect authors to explain how the study size was arrived at, which may reduce the incentive to conduct sample size calculations for observational studies [68].

However, for sample size estimations based on the clinical endpoints in the RWE studies, sample size for different study designs can have following considerations:

- When there is a given sample size or if a sample size was not specified in the protocol, it's advisable to provide power considerations for the primary analysis of the observational study to limit random errors. The power considerations necessitate a definition of a minimally important difference or intervention effect in the presence of a given sample size. Any power calculation provides the chance of a type-II error (false negative findings), while a detectable difference may be clinically more informative. For example, it shows the minimal relative risk that can be detected with the specified power and sample size given a type I error probability α [66].
- For cross, sectional studies, prospective observational studies and other quantitative data analysis, the sample size estimation can be done using the detailed methodologies reported by Charan & Biswas (2013) [69].

The statistical packages that can be generally used for analysis of real-world data include SAS, STATA, SPSS, and R.

8 QUALITY

Across the healthcare ecosystem, there are concerns over wider adoption of RWE in regulatory and reimbursement decision-making. Critics are concerned that researchers will be disincentivized from conducting RCTs and healthcare decision-makers could be forced to rely on 'inferior' evidence. Several high-profile 'disasters,' including recent retractions of a COVID-19 RWE study from major journals, have solidified the concern that RWE could lead to inaccurate results and poor patient outcomes. Critics also fear that, if allowed to do so, industry will prefer RWE instead of RCTs because RWE is cheaper. Critics thus propose continued adherence to the current paradigm of traditional evidence hierarchies, which display RCTs at the pinnacle and non-randomized studies as inferior. The lack of a gold standard in defining and creating decision-quality RWE further contributes to variability in RWE study quality, which in turn casts doubt on the validity of RWE and fuels skepticism [70].

Quality assurance of data, data collection/ registry procedures, and computerized systems are essential to provide confidence that the design, conduct, and analysis of the registry could protect against bias (systematic error) and errors in inference, that is, erroneous conclusions drawn from a study. For RWE studies – especially for registries – focus on external validity, internal validity, and analysis and reporting is essential [2].

Following are some of the guidelines which can be referred to while ensuring the quality of RWE studies:

1) **The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement:** These reporting guidelines are provided for Observational studies in epidemiology (cohort, case-control studies, cross-sectional studies). STROBE stands for an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies. The STROBE Statement is being endorsed by a growing number of biomedical journals [71].
2) **Good ReseArch for Comparative Effectiveness (GRACE):** The GRACE principles have been developed to help healthcare providers, researchers, journal readers, and editors evaluate the quality inherent in observational research studies of comparative effectiveness. The GRACE principles can be used to guide the design and evaluation of studies that are based on new data collection, use existing data, and are consistent with good pharmacoepidemiologic practice and the Agency for Healthcare Research and Quality's handbook on Registries for Evaluating Patient Outcomes [72].
3) **FDA's "Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets":** This guidance describes best practices pertaining to conducting and reporting on pharmacoepidemiologic safety studies that use electronic healthcare data, which include administrative claims data and EMR data. The

guidance includes recommendations for documenting the design, analysis, and results of pharmacoepidemiologic safety studies [73]

4) D**eveloping a Protocol for Observational Comparative Effectiveness Research: A User's Guide.** This Observational CER User's Guide serves as a resource for investigators and stakeholders when designing observational comparative effectiveness research (CER) studies, particularly those with findings that are intended to translate into decisions or actions. The User's Guide provides principles for designing research that will inform health care decisions of patients and other stakeholders. Furthermore, it serves as a reference for increasing the transparency of the methods used in a study and standardizing the review of protocols through checklists provided in every chapter [74].
5) **Guidelines for Good Pharmacoepidemiology Practices (GPP):** The GPP propose essential practices and procedures that should be considered to help ensure the quality and integrity of pharmacoepidemiologic research, and to provide adequate documentation of research methods and results. The GPP address the following areas [75]:
    - Protocol Development
    - Responsibilities, Personnel, Facilities, Resource Commitment, and Contractors
    - Study Conduct
    - Communication
    - Adverse Event Reporting
    - Archiving.
6) **Guidance for the format and content of the protocol of non-interventional post-authorisation safety studies by European Medicines Agency:** This document provides guidance for drafting the study protocols for non-interventional PASS in order to support consistency of the presentation and information provided [76].

9 DATA PRIVACY AND SECURITY

Health data requires the highest set of privacy and security safeguards. Systems must be put in place to ensure patient privacy and data security, along with a robust consent framework. Above all, there is a need to adopt a citizen-first approach when sharing health data [77]. Real world evidence studies use RWD which are generated from patient health status and/ or the delivery of health care routinely. These are collected from a variety of sources [9]. Since these studies use patient data, applicable regulations and guidelines should be followed.

9.1 Informed consent

Data privacy and security is an integral part of RWE studies and it has to be ensured that all the applicable guidelines and regulations are followed. For example, informed consent is a key tenet in medicine and is often understood as the explicit documented approval given by a patient to receive medical interventions after having reflected on related benefits and harms. The seeking of consent to collect and use patients' data—including from their medical records, radiological images and tissue samples—has historically been less explicit [78].

In most primary care settings in India, general practitioners seldom maintain any records, and consent is not sought when they do. Community health workers routinely collect large volumes of data without explicit consent or explanation about how the data will be used. In modern hospitals, if consent is sought for the collection or use of data, it is documented during patient registration or at the bedside just prior to interventions. According to the more recent telemedicine guidelines, if a patient initiates a telemedicine consultation, her consent is implied and not required to be explicitly sought. This is not true 'autonomous authorisation'. The power hierarchy operating in such interactions likely impedes true autonomous decision-making and is particularly exacerbated when services are sought by individuals already discriminated against due to gender, caste or class [78]. Health data are also increasingly exchanged across services such as wearables, applications and some point of-care devices that are governed by weak data protection regulations. The language, length and complexity of consent documents accessed through small screens on mobile devices or wearables with little or no true choices have rendered them irrelevant, opaque, non-comparable and inflexible [78].

In light of the above, it is critical that it is understood that the owner of the patient's data is the patient himself or herself and all data privacy and security measures are taken while conducting an RWE study.

In Table 4 below, existing framework for data protection in India is provided. Some important framework and its applicability are discussed in this paper [78].

Table 4: Existing framework for data protection in India

Document	Details	Type	Nature
Puttaswamy versus Union of India	Judgement of the Supreme Court of India affirming the right to privacy of all individuals under the Indian Constitution	Law	Binding
Information Technology Act, 2000	Prescribes security practices for the protection of personal data. Requires that consent must be sought before the collection of any sensitive personal data	Law	Binding and enforceable
HIV/ AIDS Act 2017, Mental Healthcare Act, 2017, Transplantation of Human Organs and Tissues Act, 1994	Sector-specific laws that govern data related to the disease area. The requirements may be different from those under the Information Technology (IT) Act	Law	Binding and enforceable
Personal Data Protection Bill, 2019	Proposed law that updates the IT Act and protects all personal data, establishes a data protection regulator and prescribes penalties for violations of these rules	Bill; pending in parliament	Unenforceable till passed as law
Data Empowerment and Protection Architecture	Framework for data management and security issued by NITI Aayog, a government think-tank	Draft report	Voluntary
National Digital Health Blueprint, NDHM Health Data Management Policy, NDHM strategy overview	Lays out the architectural framework for the digital health infrastructure under the NDHM	Government reports	Voluntary
Report by the committee of experts on Non-Personal Data Governance Framework	This committee of experts was constituted by the Ministry of Electronics and IT to propose a governance framework for non-personal data. It has released a draft report for public comments (July 2020).	Draft government report	Recommendations to the government

9.2 Information Technology Act, 2000 ('IT Act') and Information Technology (Amendment) Act 2008 (the '2008 Act')

Until the Information Technology Act, 2000 ('IT Act'), as originally enacted, India lacked a central statutory framework for data protection. It provided for civil and criminal actions in cases of gaining unauthorized access and downloading or extracting data stored in computer systems or networks or tampering with computer source code, hacking with an intent to cause damage, and

breach of confidentiality and privacy. However, it did not contain any positive data protection obligations [79].

The IT Act was amended in 2008 by the Information Technology (Amendment) Act 2008 (the '2008 Act'). This added a new section 43A which expressly recognized compensatory relief for a body corporate's failure to protect sensitive personal data or information of a person. 'Body corporate' is defined as any company and includes a firm sole proprietorship, or other association of individuals engaged in commercial or professional activities [79].

The most important features and guidance provided are:

1. The 2008 Act also provides that 'reasonable security practices and procedures' means security practices and procedures designed to protect such information from unauthorized access, damage, use, modification, disclosure, or impairment, as may be specified by any of the following ways:
    - in an agreement between the parties;
    - in any law for the time being in force;
    - in the absence of the foregoing, such reasonable security practices and procedures, as may be prescribed by the Central Government in consultation with such professional bodies or associations as it may deem fit [79].
2. Definition of sensitive personal data or information:

    The following types of information are identified as sensitive personal data or information:
    - passwords;
    - financial information such as bank account or credit card or debit card or other payment instrument details;
    - physical, physiological, and mental health conditions;
    - sexual orientation;
    - medical records and history;
    - biometric information [79]

    Any information that is freely available or accessible in the public domain or furnished under the Right to Information Act 2005 or any other law for the time being in force is not to be regarded as sensitive personal data or information for the purposes of these rules [79].

3. Disclosure of information to third parties:

    Disclosure of sensitive personal data or information by a body corporate to any third party requires prior permission from its provider, who has provided such information under lawful contract or otherwise, unless that disclosure has been agreed in the contract between the body corporate and provider of information, or where the disclosure is

necessary for compliance of a legal obligation. Prior consent from the provider of information may not be required in two circumstances:

- where the information is shared with Government agencies mandated under the law to obtain information including sensitive personal data or information for the purposes of verification of identity, or for prevention, detection, investigation including cyber incidents, prosecution, and punishment of offences
- where the information must be disclosed to any third party by an order under the law for the time being in force.

The third party receiving the sensitive personal data or information from body corporate or any person on its behalf is under an obligation not to disclose it further [79].

In view of these rules, companies outsourcing their operations to India should carefully examine contracts with the Indian offshore partners to ensure compliance with them and put into place the prescribed procedures [79].

9.3 EHR standards

With the Government of India's 'Digital Health Mission', there has been a lot of advancements in the conglomerations of Healthcare and IT industry. This movement has led to the development of new methods and tools in maintaining the patient data in the digital form. The Electronic Health Record (EHR) is one such solution to support the healthcare facility, irrespective of levels and sizes to improve patient care by enabling functions that other types of records cannot deliver. The major requirements in healthcare facility is to use interoperability and standardization technique to enable easy sharing and exchange of healthcare data between the various levels. The main foundation for the interoperability is the standard terminology, which improves the effective communication between the two healthcare users. Government of India has taken the initiative and formulated and published an EHR standard in September 2013 and consequently revised and published the next version on 31 December 2016 (2016). For EHR standards, Ministry of Health and Family Welfare (MoHFW) suggest the healthcare facility to use the following standards namely Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) (2016), International Classification of Diseases (ICD 11) (2016), Logical Observation Identifiers Names and Codes (LOINC) (2016), and National Drug Code(NDC) (2016) [80].

As per the EHR standards released by the Ministry of Health and Family Welfare, Government of India, to create an Electronic Health Record (EHR) of an individual, it is essential that all clinical health records created by the various health care providers that a person visits during his/ her lifetime be stored in a central clinical data repository and also be shareable through the use of interoperable standards. Adequate safeguards to ensure data privacy and security must strictly be included at all the times since the patient's medical data are so sensitive and to be protected. Patients must have the privilege to verify the accuracy of their health data and gain access. The EHR standards of India emphasize on patient as the authorized owner of his health data. The

standards aim to develop a system which would allow one to create, store, transmit or receive electronically using reliable media for data storage and transfer. EHRs can bring a patient's complete health information together for better medical decisions [81].

9.4 Personal Data Protection (PDP) bill 2019

In India, the dire need for data protection framework was perceived subsequent to the case of Justice K.S. Puttaswamy (Retd.) & Anr v Union of India & Ors1 ("Privacy Judgment"). On this account, a committee was formed in July 2017 by the ministry of electronics and information technology which consisted of ten experts leaded by a former Supreme Court judge, Justice B.N Srikrishna. On 27th July 2018, the ten-membered committee submitted its report on data protection law which was revised by the government. Subsequently, on 11th December 2019, the revised version of the 2018 Draft Bill known as the Personal Data Protection Bill, 2019 ("PDP Bill") was presented before the Indian Parliament. The principles of the PDP bill are tantamount to that of GDPR (General Data Protection Regulation) broadly as the PDP has incorporated certain aspects of GDPR in the bill. However, there are differences and variance between the two legislations [82].

The Personal Data Protection Bill, 2019 ("PDP Bill" or "The Bill") was introduced in the Lok Sabha on December 11, 2019. The PDP Bill revises an earlier version of the Bill drafted by a Committee of Experts under the Chairmanship of Justice B.N. Srikrishna in 2018 ("2018 Bill"). The Bill amends the Information Technology Act, 2000 ("IT Act") to repeal the provisions (Sections 43A and 87) that currently deal with data protection. Thus, if passed, it will replace the existing data protection framework under the IT Act and under the Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011 [83].

Like GDPR, this draft bill also deals with protection of personal information and data. Data under this Bill has been bifurcated into three parts:

1. Personal Data: Personal data includes any information related to a natural person known as Data Principal, which could potentially reveal the identity of the individual.
2. Sensitive Personal Data: Sensitive personal data includes passwords, health and financial data, biometric and genetic data, official identifier, information revealing about one's sex life, sexual orientation, transgender and intersex status and caste tribe.
3. Critical Personal Data: The definition of Critical Personal Data has not been stipulated in the Bill. It will be notified by the Central Government [82]

The Bill creates a consent-based framework for processing of Personal data. Personal data cannot be processed without the consent of the individual to whom the data relates to, referred to as "Data Principal". This consent has to be sought at the beginning of the process. The consent of the Data Principal must be free, informed, specific, clear and capable of being withdrawn. To ease the process of consent, the Bill provides for registration of Consent Managers, through which the Data Principal may give or withdraw their consent. At the time of collection of personal data, which has also been classified under data processing, the Data Fiduciary is required to give a

notice to the Data Principal. Data Fiduciary means any person, including the State, a company, any juristic entity or any individual who alone or in conjunction with others determines the purpose and means of processing of personal data [83].

The PDP Bill also imposes obligations on Data Fiduciaries to regulate the purpose and methods by which personal data can be processed and collected. Personal data can be processed only for a specific, clear and lawful purpose, and such processing must be carried out in a fair and reasonable manner, while ensuring the privacy of the data principal. Importantly, the processing must be for the purpose consented to by the Data Principal or which is incidental to or connected with such purpose. The collection of personal data must also be limited to the extent necessary for the purposes of processing. By ensuring that the purpose for which the data is being processed is under the control of the Data Principal, the PDP Bill makes it certain that the Data Principal will be aware of the metadata that will be generated from their personal data. Furthermore, Section 17(1)(b) of the PDP Bill gives the right to the Data Principal to access a summary of the data being processed. However, the Bill fails to give any rights to the Data Principal to access the processed data or analysis thereof. Lee – Berners expresses his concern about this in the Contract, as this metadata could be utilized further without the Data Principal being aware [83].

Today a bigger threat is not just the big data companies having access to private and confidential data but also the Government having access to it. Clause 2 of the Third Principle of the Contract is concerned with the government seeking our private communications and private data. It seeks to establish due process based on international human rights norms, which do not weaken the security of service providers. Clause 3 of the Third Principle of the Contract follows on from this, requesting the government to support and monitor online data privacy rights in its territory, by regulating the data companies and limiting their own data collection to what is necessary to achieve specific public interests [83].

Section 33 of the PDP Bill states that sensitive personal data may be kept outside India, as long as a copy of the same stays within the country. However, Critical Personal Data has to be processed and stored within India only. This provision permits personal data to be kept outside India whilst not maintaining a copy of the same within the territorial jurisdiction of India [83].

Personal Data Protection Bill is applicable to the corporations and industries that:

- Processes the data which is gathered, shared or stored within Indian Territory.
- Processes personal data associated with any business carried on within Indian Territory.
- Processes personal data of any natural person within the territory of India.

The bill also deals with:

- Processing of personal data by the State, any Indian company or any Indian citizen or persons incorporated under the Indian law.

- Processing of personal data by data fiduciaries or data processors not present within the territory of India if:
    - Such processing is linked with any business carried on within Indian Territory
    - Such processing is associated with any activity involving profiling of natural persons within Indian Territory [82]

Data Fiduciary and Data Processor could be any person, company, juristic entity or state:

- Data Principal: The natural person who is the subject of personal data is known as the natural person.
- Data Fiduciary: An individual who alone or in conjunction with others, ascertains the objective and mode of processing of personal data is the data fiduciary.
- Data Processor: An individual who is not an employee of the fiduciary and processes data on behalf of the fiduciary is the data processor.

9.5 Health Data management policy (2020):

The Ministry of Health and Family Welfare ("MoHFW") is responsible for conceiving the idea of the National Digital Health Mission ("NDHM"). This visionary project of the Government of India, stemming from the National Health Policy, 2017 ("National Health Policy") intends to digitise the entire healthcare ecosystem of India. This would be done by creating digital health records, and creating and maintaining registries for healthcare professionals and health facilities in order to ensure a smooth interoperable framework for the multiple partners associated with healthcare delivery to individuals in India. The National Digital Health Blueprint, 2019 ("Blueprint") recommends that a federated architecture be adopted, instead of a centralised architecture, for the management of digital health data to ensure interoperability, technological flexibility and independence across the National Digital Health Ecosystem ("NDHE") [77].

This Health Data Management Policy ("Policy") is the first step in realising the NDHM's guiding principle of "Security and Privacy by Design" for the protection of individuals'/ data principal's personal digital health data privacy. It acts as a guidance document across the NDHE and sets out the minimum standard for data privacy protection that should be followed across the board in order to ensure compliance with relevant and applicable laws, rules and regulations. This Policy will be dynamic in nature and may be revised from time to time as may be required. Necessary guidelines may also be issued for the implementation of the NDHM [77].

NDHM defines the below terms which are critical for RWE studies in data privacy [77]

1. "Anonymisation" in relation to personal data, means such irreversible process of transforming or converting personal data to a form in which a data principal cannot be identified through any means reasonably likely to be used to identify such data principal;
2. "Child" means a natural person/ individual who has not completed eighteen years of age;
3. "Consent" means the consent referred to in Clause 9 of this Policy; (given below)

4. “Consent artifact” means a machine-readable document that specifies the parameters and scope of data sharing and access that a data principal consents to in any personal data sharing transaction;
5. “Consent manager” means an electronic system that interacts with the data principal and obtains consent from him/ her for any intended access to personal data;
6. “Data” means and includes a representation of information, facts, concepts, opinions, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automated means;
7. “Data fiduciary” means any person, including the State, a company, any juristic entity or any individual who alone, or in conjunction with others, determines the purpose and means of processing of personal data. For the purpose of this Policy, data fiduciaries would include Health Information Providers and Health Information Users if such entities are determining the purpose and means of processing of personal data;
8. “Data principal” means the natural person/ individual to whom the personal data relates;
9. “Data processor” means any person, including the State, a company, any juristic entity or any individual, who processes personal data on behalf of a data fiduciary;
10. “De-identification” means the process by which a data fiduciary or data processor may remove or mask identifiers from personal data, or replace them with such other fictitious name or code that is unique to a data principal but does not, on its own, directly identify the data principal;
11. “Electronic health records” or “EHR” are one or more repositories, physically or virtually integrated, of data in digital form, relevant to the wellness, health and healthcare of an individual, capable of being stored and communicated securely and of being accessible by multiple authorized users (such as healthcare professionals or health facilities), represented according to a standardized or commonly agreed logical information model. Essentially, an EHR is a collection of various medical records that get generated during any clinical encounter or events;
12. “Electronic medical records” or “EMR” refers to a repository of records that is stored and used by the HIP generating such records to support patient diagnosis and treatment. EMR may be considered as a special case of EHR, limited in scope to the medical domain or is focused on the medical transaction;
13. “Health facility” refers to health facilities across the country and includes hospitals, clinics, diagnostic centres, health and wellness centres, mobile vans, ambulances and pharmacies;
14. “Personal data” means data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute or any other feature of the identity of such natural person, whether online or offline, or any

combination of such features with any other information. For the purpose of this Policy, personal data would include Personal Health Identifier;

15. "Pseudonymisation" means a data management and de-identification procedure by which personally identifiable information fields within a data record are replaced by one or more artificial identifiers, or pseudonyms;

The NDMP provides Facility ID which are unique ID allocated to each health facility. In addition, "Health ID" refers to the Identification Number or Identifier allocated to a data principal in accordance with Chapter IV of this Policy. Health ID Provider are also given to persons such as data fiduciaries, HIPs or HIUs which have been authorised by the NDHM to issue Health IDs [77].

9.6 Consent in relation to collection and processing of personal data:

Data fiduciaries can collect or process personal data only with the consent of the data principal. It is the responsibility of the data fiduciary to ensure that the consent given by the data principal is valid. The consent of the data principal will be considered valid only if it is:

- Free, having regard to whether it complies with the standards set out under Section 14 of the Indian Contract Act, 1872;
- Informed, having regard to whether the data principal has been provided with the necessary information by way of notice, as set out in Clause 10 of this Policy, the scope of consent in respect of the purpose of processing;
- Specific, where the data principal can give consent for the processing of personal data for a particular purpose;
- Clearly given; and
- Capable of being withdrawn at any time, having regard to whether the ease of such withdrawal is comparable to the ease with which consent may be given.

The purposes for collection or processing of personal data shall be limited to those which may be specified by the NDHM and such purposes will be related to the health of an individual or may be such other incidental purposes which a data principal can reasonably expect, having regard to the purpose and the context and circumstances in which the personal data was collected or processed. In addition to the conditions mentioned in above, the consent of a data principal in respect of collecting or processing any sensitive personal data will be obtained only after informing her/ him the purpose of, or operations in, processing which are likely to cause significant harm to the data principal [77].

The NDMP provides guidance for [77]:

1. Method of obtaining consent
2. Processing personal data pertaining to a child
3. Processing personal data of data principals who are seriously ill or mentally incapacitated, or in response to a medical emergency involving a threat to the life or a severe threat to the health of the data principal

An HIU shall follow the principle of data minimisation and shall obtain the consent of the data principal only for such personal data that is necessary for the purposes for which such consent is being sought.

Sharing of personal data and obligations of entities with whom personal data is shared: important points under this are:

- Any personal data processed by a data fiduciary may be shared with an HIU in response to a request made by such HIU for personal data pertaining to the data principal, only where consent of the data principal is obtained.
- A data fiduciary shall maintain a record of all consent obtained under this Policy, pursuant to which personal data has been shared by such fiduciary under this Policy in a manner that enables the audit and review of such data sharing.
- In addition to the obligations set out in Chapter V, an HIU shall ensure that any personal data under this Policy: (a) shall not be used by the HIU for any purpose other than what was specified to the data principal at the time of obtaining his/ her consent (b) shall not be disclosed further without obtaining the consent of the data principal for such disclosure

9.7 What can be done:

9.7.1 Privacy by design

Privacy by design (PbD), a systems engineering approach first developed by Cavoukian in 1995, calls for proactive privacy preserving design choices embedded throughout the process life cycle. Since the advent of EMRs, experts have recognised the need for embedding technological safeguards to protect privacy and prevent data breaches. Advances in data science help address several of the aforementioned limitations, by either manipulating the data through strategies like minimisation, separation or abstraction or regulating the process by defining conditions for control and notification. In many settings in India, personal data can often be easily accessed by people who do not need such access; for example, clinic-based facilitators that liaise with state or private insurance companies, insurance agents themselves and in the public sector, administrative officials. There is little recognition that such access, however unintentional or inadvertent, is unethical, and will very soon be illegal. The NDHM strategy calls for PbD tools without providing greater detail. We have described below the dominant tools in current use that apply PbD principles to address gaps in health data protection. These examples are meant to be illustrative and are not exhaustive [78].

9.7.2 Data minimisation

When health data are collected, either through clinical operations or during research, there is temptation to collect more and not less, given the opportunity costs associated with collecting these data. This results in exhaustive data sets archived in the public and private health sector that pose significant privacy risks. Restricting data collection to the essentials has in fact been

demonstrated to declutter and improve the user-interface, and consequently, user-experience and compliance, while reducing privacy risks. While the NDHM espouses data minimisation, existing legacy digital public health systems continue to collect vast amounts of redundant data on millions of beneficiaries, without demonstrable justification [78].

9.7.3 Role-based access

Role-based access is a standard feature in most advanced EMRs. Open source tools like Dataverse provide scientists differential access to research databases as well. Multi-authority attribute-based encryption schemes allow role-based models to scale by allowing access to users based on a set of attributes, rather than on individual identities. For example, by virtue of being a verified clinician (regardless of who), physicians are generally able to look up most medical records at their institution easily; by virtue of being a public health administrator (regardless of who), officers should have no access to personal health information; and by virtue of being a research laboratory, the team would have access to authorised de-identified data, provided third-party regulators can affirm the veracity of each of their attributes (clinician, administrator, researcher). The Account Aggregator, a similar consent management framework already in play in India's fintech ecosystem, lends itself to such selective, verifiable, pre-authenticated access as has been proposed at the backbone for the NDHM. Since user consent can be sought asynchronously (prior to actual data processing), this model somewhat mitigates inadvertent coercion associated with point-of-care consent seeking. The NDHM seeks to verify attributes by developing and maintaining 'registries' of providers [78].

9.7.4 User preference

The General Data Protection Regulation in the European Union facilitates data access by requiring companies to provide a consent management platform to give users more control over their data, by selecting from a menu of data-use options. In India, the Data Empowerment and Protection Architecture and the NDHM seek to empower users by allowing them to place revocable time and purpose limitations on the use of their data— the sorts of choices that would be extremely beneficial to patients. In theory, patients would control who accesses their data at all times, would receive notification of third party access (whether authorised or not), or be able to revoke access at will, when permitted by law. Others have elaborated on the idea by allowing data principals to opt into certain 'data trusts' or stewards with pre-negotiated access controls, where general attributes can be used to guide future data sharing: for example, a patient may elect to always allow healthcare providers to access her data but always deny access to pharmaceutical companies regardless of the identifiability of the data. This approach would entail data principals communicating their preferences to the consent manager to accordingly direct data toward select categories of data processers; for example, to clinical health information users, and say, public research agencies like the ICMR, but not to pharmaceutical companies. The asynchronous and one-time (but revocable and changeable) nature of the process—made possible by the consent manager framework—may allow users to make a more informed and coercion-free choice, if citizens are encouraged to actively enroll in the system prior to clinical care [78].

9.7.5 Differential privacy

The current NDHM guidelines require that all health information processors make aggregated data available. Not only are aggregation and anonymisation inadequate for protecting privacy for the reasons described above, but many aspects of clinical and population health will require non-anonymised, high resolution data to actually be useable and useful. The NDHM's Health Data Management Policy prohibits inadvertent unforeseen re-identification while processing data. Differential privacy (DP) seeks to balance such access to rich data while preserving privacy. It achieves this balance by differentially introducing 'statistical noise' in the data set, depending on what is being queried and by whom, thus combining the aforementioned approaches. The 'noise' masks the contribution of each individual data point without significantly impacting the accuracy of the analysis. Moreover, the amount of information revealed from each query is calculated and deducted from an overall privacy budget to halt additional queries when personal privacy may be compromised. If effective, this approach will help alleviate some of the concerns about combining large data sets; its utility in the clinical setting is yet to be determined. There is precedent for DP as a model for collaborative research. Open source platforms like OpenDP are likely to accelerate use of the application of DP across disciplines. DP may however lead to noisy aggregates with poor utility for analytical tasks in public health. Given the nascency of DP applications, it is premature to assess utility based on field-impact [78].

9.7.6 Regulation

The jurisprudence on privacy is rapidly evolving in India, and includes a landmark judgement of the Supreme Court affirming the right to privacy. The PDP Bill seeks to regulate the collection and transfer of all personal data, including health information. The law requires consent from the data principal before processing their personal data, and because health data are considered 'sensitive' by the law, the data principal [78].

10 STUDY REPORT

Real world evidence, generated from sources of RWD in various health care settings and geographic locations, present opportunities for innovative, efficient, and cost-effective research to inform decisions about the clinical effectiveness and safety of medical products and interventions in clinical medicine, health services, and public health [33]. However, there may be barriers to the use of these studies due to the limited number of accepted principles for their evaluation and interpretation [77, 33]. Regulators are increasingly calling for high levels of transparency and reproducibility as an integral part of the science of RWE [84]. Transparency is based on openness, communication, and disclosure of information, whilst respecting the protection of both personal data and commercially confidential information.

The importance of achieving consistently reproducible research is recognized in many reporting guidelines. Reporting guidelines have been developed to guide reporting for a range of study designs and contexts and are associated with improved quality of reporting [85]. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement was developed to enhance the transparency of reporting of observational research. Most research conducted using routinely collected data is observational in design, and therefore, the STROBE guidelines are relevant and applicable. Whilst the STROBE statement is designed to apply to all observational studies, specific issues related to reporting research using routinely collected data are not fully addressed. Thus, the REporting of studies Conducted using Observational Routinely collected Data (RECORD) initiative was established as an international collaborative process and an expansion of STROBE to explore and address specific reporting issues relevant to research using routinely collected health data [85]. It may be useful to report the results of observational studies of comparative effectiveness in the context of how well they support existing clinical trial data. Reporting observational comparative effectiveness studies may contribute to a better clinical and biological understanding of the disease, either by confirmation in a more targeted RCT or through advances in basic science. The Good ReseArch for Comparative Effectiveness (GRACE) principles describe a hierarchy of evidence for observational research on comparative effectiveness that can be used by decision-makers, as well as key elements of good practice including defining research questions and methods a priori; collecting valid, clinically relevant data; analyzing, interpreting and reporting data, including sensitivity analyses and alternative explanations for findings; and conducting these studies in accordance with accepted good practices [86].

Incomplete and inadequate reporting of research hampers the assessment of strengths and weaknesses of studies being reported. If observational studies based on RWD are to be accepted as valuable sources of evidence, complete reporting is required. The outcome of a study shall always be presented in an objective and truthful manner, providing a comprehensive and accurate description of the findings. A clear summary of the main results of the study, whether positive or negative shall always be made available. For the content of the report(s), it is recommended to follow the Guidelines for Good Pharmacoepidemiology Practices (GPP) [87] of

the International Society of Pharmacoepidemiology (ISPE) and the STROBE [89] and RECORD statements[85] for study reports.

Key Elements to be included in the study report are [85, 86, 87, 88]:

1. **Title and Abstract:**

Title page should include (whatever is applicable) the study's design with a commonly used term, purpose of the study, as stated in the protocol, the type of routine data (e.g., health administrative data, other administrative data, disease registries, primary care databases, electronic health record data, and population registries) used, the geographic region to define the study population (e.g., nation, state, province, and region), time frame with dates on which the study was initiated and completed, and linkage between databases (if it was conducted). The names, titles, degrees, addresses, and affiliations of the principal investigator and all co-investigators and name and address of the Sponsor should be provided in the Title page. Abstract should be an informative and balanced summary of what was done and what was observed.

2. **Introduction:**

The scientific background and rationale for the investigation being reported is explained in this section. Stating the specific research objectives is essential for replication and translation of any observational research. For studies using routinely collected data, it should be further clarified whether the analyses were exploratory with the purpose of finding new relationships in the data (examples are data mining or hypothesis-generating studies) or confirmatory with the purpose of testing one or more hypotheses. This section should indicate whether the hypotheses were generated before or after data analysis. It should be clearly stated whether there is a study protocol and how this can be accessed and if the study was registered in a publicly accessible study registry. A clear description of a study's objectives is essential. It is insufficient to simply label a study as descriptive without clarifying whether it aims to generate or examine a hypothesis.

3. **Methodology:**

This should include design, population, disease/ condition, comparators, variables, and setting with the perspective of particular decision-makers in mind (e.g., payer or provider).

- *Design*: A research plan should be developed before starting the study. The study plan should include clinically meaningful outcome measures that will assist patients and health professionals with treatment decisions or policymakers with decisions about allocations of resources. The study plan should be sufficiently detailed to allow replication of methodology.
- *Disease/ Condition*: This includes diagnostic certainty, severity, time since diagnosis, significant co-morbidities, treatment history, etc.

- *Population*: Population characteristics to consider include demographics such as age and gender, nationality, ethnicity; risk factors such as average blood pressure, cholesterol levels, and body mass index; behaviors such as smoking; disease history, onset, stage, and severity of the condition; past and current treatments for the condition; and clinical issues such as co-morbidities.

- *Comparators*: It is frequently unrealistic that ALL interventions that could be considered to treat a disease or condition be included in the analysis; however, omitting key interventions that represent standards of care introduces the potential for bias and uncertainty. Comparisons to a number of real-world alternatives are generally preferable to a single comparator. To what extent the treatment and its therapeutic alternatives are already in use in the target population in sufficient numbers should be considered for meaningful analysis and interpretation. For each comparator, consider the brand, dosage, method of delivery, duration of use, whether for a labeled indication, therapeutic alternatives currently in use, the likelihood that the necessary information will be accurately recorded and accessible, etc.

- *Variables*: All outcomes, exposures, and predictors should be clearly defined. Diagnostic criteria should be given, if applicable. The outcome measure (endpoint) should be clearly stated, clinically meaningful, and appropriate for measuring effectiveness. Outcomes such as cardiovascular events (e.g., rates of myocardial infarction or stroke), mortality, patient functioning, health-related quality of life or health status measures (e.g., scores from the SF-36 Health Survey or the EQ-5D) may be more relevant to a decision-maker than surrogate or intermediate endpoints (e.g., cholesterol levels).

  Exposure to treatment is ideally documented by evidence that patients actually took the medication or received the treatment as per routine practice. Exposure may be documented by evidence of a prescription being written, a claim being filed, and measures of medication possession.

  Misclassification of drug exposure can result from the patient's incorrect recall of dose or poor adherence or treatment compliance. Studies using data sources that track prescription fills and refills are particularly vulnerable for treatments used on an as-needed basis (e.g., migraine medications) and for treatments dispensed in liquid or inhalable forms.

- *Settings*: Factors that should be considered may include the study time frame, the payer setting, provider characteristics, or the geographic area. The setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection are described.

4. Data Collection:

How the data were collected, enrollment and coverage factors, pathways to care, quality assurance, other factors that may have affected the quality of the data and the validity of

conclusions that may be drawn from their analysis should be described for both primary collection of data, i.e., data that are collected specifically for the purposes of the study, and secondary use of data, i.e., data that were collected for other purposes (such as administrative claims data and medical records). The data source and how the data were collected is essential to document when using secondary data as this will help minimize errors in interpretation. For example, billing codes may be recorded inaccurately (e.g., coding errors), imprecisely (e.g., DRGs), inconsistently, and/ or under different constraints (e.g., one intervention might have been subject to pharmacy prescription limits, such as for migraine medicines) and may thus not be reflective of the actual clinical condition. For routine data consisting of survey results, the survey questions should be provided with the precise wording given to study participants.

5. **Statistical Methods:**

This includes how populations, interventions, and outcomes are defined; how missing data are dealt with; how outliers are dealt with; which analytic approaches were taken, and to what extent unmeasured confounders may influence the results. Whether the study included individual-level, institutional-level, or other data linkage across two or more databases should be mentioned. Linkage techniques and methods used to evaluate linkage quality should be provided. The description of data cleaning methods at different stages of the study should include those used to screen for erroneous and missing data, including range checks, checks for duplicate records, and handling of repeated measures. The most common and basic approaches for identifying heterogeneous treatment effects are to conduct subgroup analyses. Any methods used to examine subgroups and interactions are described. If sensitivity analyses were conducted based on different sets of codes/ algorithms, these should also be described and evaluated.

6. **Bias and Confounding factors:**

Accurate interpretation depends on understanding the extent to which bias (stemming from factors that are related both to the decision to treat and to the outcome(s) of interest) may have distorted the results.

A confounder is a factor that distorts the true relationship of the study variables of central interest by virtue of being related to the outcome of interest, but not related to the study question and unequally distributed among the groups being compared. Stronger methods to deal with potential confounding that may occur due to lack of randomization include inception cohorts, new user designs, the use of multiple comparator groups, matching designs, and assessment of outcomes not thought to be impacted by the intervention compared.

Various types of bias to consider include:

- *Selection bias* refers to systematic differences among the groups being compared that arise from self-selection or physician-directed selection of treatments, or association of treatment assignments with other characteristics such as education, ethnicity, age, access to healthcare, etc.

- *Misclassification bias* occurs when an exposure or outcome is incorrect or missing.
- *Detection bias* applies to situations in which comparison groups are assessed at different points in time or using different methods or by assessors who may have knowledge of which treatment was used.
- *Performance bias* refers to systematic differences in care other than the intervention under study.
- *Attrition* refers to selective loss to follow-up.

7. Results:

Reporting the number of individuals screened at each stage of the selection process is important to assess the potential selection bias of participants and can provide cursory evidence that the study procedures were implemented correctly. The final analyzable sample can most easily be interpreted when a text description or a flow diagram is provided that describes the initial pool of potential subjects and the sample after each inclusion and exclusion criteria is applied. A basic summary of the observable characteristics of the study population should be provided including descriptive statistics on the mean value (and distribution where appropriate) for demographic variables, prevalence of co-morbidities, and other potential confounders reported by treatment groups.

Sufficient tables, graphs, and illustrations should be included to present the pertinent data and to reflect the analyses performed. Epidemiologic parameters (e.g., risks, rates, risk or rate differences, and risk or rate ratios) are the most typical epidemiologic measures to report. Both unadjusted and adjusted results should be presented. Effect measures should not be described as “significant” or “not significant.” Precision of estimates should be quantified using confidence intervals. Confidence intervals communicate both the strength of the relationship and the precision of the measure and are therefore more informative than point estimates accompanied by p-values. Absolute measures of effect included differences in proportions, means, rates, number-needed-to-harm (NNH), number-needed-to-treat-to-harm (NNT_H) and number-needed-to-treat (NNT) and should be reported for a meaningful time period. Relative measures of effect are rate ratios, proportions, or other measures and include odds ratios (ORs), incidence rate ratios, relative risks, and hazard ratios (HR).

Statistical techniques like propensity score methods and instrument variable methods used to adjust for multiple analyses of the same data, reporting of unadjusted estimates of treatment effects should be reported. Describe how the missing data were handled. The study must report confounder-adjusted estimates if they are attempting to make any inference regarding the effects from treatment. Unadjusted estimates should also be reported to allow for comparison with the adjusted results.

8. Discussion:

The implications of using data that were not created or collected to answer the specific research question(s) should be discussed. Any major deviations from the protocol, that may potentially impact the study population or add bias should be addressed. Discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported should be included. Another important potential limitation is changes in coding practices or eligibility criteria resulting from a change in the composition of the database population, study population, or both over time.

To aid interpretation of research, a thorough review of literature should be undertaken to compare the findings to all known previous findings exploring the same or similar objectives. The report should provide plausible explanations for disparate findings and identify methodological differences or advance a theoretical or biologic rationale for the differences. The report should provide plausible explanations that have led to findings that are different in direction or magnitude. A statement of the conclusions drawn from the analyses of the data should be added.

11 RWE COUNCIL WORKING GROUP

RWE Topic	Author
Introduction	Vaishnavi Sivasankar, Manager-Medical Governance & Regional medical Lead-Established products (APAC), Bayer Zydus Pharma Private Limited
Real-world evidence study designs	Bruno Jolain MD, Chief Medical Officer (CMO) Roche Products (India) Pvt. Ltd
Conduct of Real-world evidence studies	Kavitha Gurram, Director & Head of Project Management, Real-World Evidence Solutions IQVIA
Informed consent	Dr Poonam Sule, Senior Scientist, Clinical Development, PHC R&D, Procter and Gamble Health Limited
Regulatory requirements - Real-world evidence studies – India	Dr. Nitin Maksane, Global Medical Affairs Manager, GSK
Analysis: Statistical considerations for Real-world evidence studies	Mahendra Kumar Rai, Senior Director HEOR & RWE, EVERSANA Asia Pte Ltd, Singapore.
Quality	Dr Deepa Chodankar, Heading GENESIS unit, Sanofi
Data privacy and security	Dr Deepa Chodankar, Heading GENESIS unit, Sanofi
Study report	Shruti MP, Associate Director, Real World Data Solutions, Parexel

11.1 Conflict of interests/disclosures

None.

11.2 Acknowledgements

We would like to thank Dr Arun Bhatt, (Consultant, Clinical Research & Drug Development) for his guidance for this paper and Saranya Vellangallor Subramanian (Senior PM, Medical Writing, Zifo RnD Solutions) for writing and editorial support.

12 REFERENCES

1. "Food and Drug Administration. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. Available from: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf. [Last accessed on 2017 Sep 03]."
2. Bhatt A. Conducting real-world evidence studies in India. Perspect Clin Res. 2019 Apr-Jun;10(2):51-56. doi: 10.4103/picr.PICR_8_19. PMID: 31008069
3. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac-Dabic D, Marks PW, Robb MA, Shuren J, Temple R, Woodcock J, Yue LQ, Califf RM. Real-World Evidence - What Is It and What Can It Tell Us? N Engl J Med. 2016 Dec 8;375(23):2293-2297. doi: 10.1056/NEJMsb1609216. PMID: 27959688.
4. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. Br J Cancer. 2014 Feb 4;110(3):551-5. doi: 10.1038/bjc.2013.725. Epub 2014 Jan 14. PMID: 24495873;
5. Bhatt A. Real-world data in COVID-19 pandemic: An essential unmet health-care need. Perspect Clin Res. 2020 Jul-Sep;11(3):103-105. doi: 10.4103/picr.PICR_168_20. Epub 2020 Jul 6. PMID: 33033697
6. US Food and Drug Administration Coronavirus (COVID-19) Update: FDA Collaborations Promote Rigorous Analyses of Real-World Data to Inform Pandemic Response. [Last accessed on 2021 Oct 25]. Available from: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-collaborations-promote-rigorous-analyses-real-world-data-inform
7. Williamson E, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Open safely: Factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. medRxiv Preprint. 2020 [Doi: 10.1101/2020.05.06.20092999]
8. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, Li Y, Guan W, Sang L, Lu J, Xu Y, Chen G, Guo H, Guo J, Chen Z, Zhao Y, Li S, Zhang N, Zhong N, He J; China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med. 2020 Aug 1;180(8):1081-1089. doi: 10.1001/jamainternmed.2020.2033. PMID: 32396163
9. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence
10. Physicians see big promise in use of real-world data, evidence. American Medical Association. https://www.ama-assn.org/practice-management/digital/physicians-see-

big-promise-use-real-world-data-evidence. Published 2020. Accessed September 21, 2020."

11. "Bhatt A. Artificial intelligence in managing clinical trial design and conduct: Man and machine still on the learning curve? Perspect Clin Res. 2021 Jan-Mar;12(1):1-3. doi: 10.4103/picr.PICR_312_20. Epub 2021 Jan 19. PMID: 33816201; PMCID: PMC8011519.
12. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017 Jun 21;2(4):230-243. doi: 10.1136/svn-2017-000101. PMID: 29507784
13. Harrer S, Shah P, Antony B, Hu J. Artificial Intelligence for Clinical Trial Design. Trends Pharmacol Sci. 2019 Aug;40(8):577-591. doi: 10.1016/j.tips.2019.05.005. Epub 2019 Jul 17. PMID: 31326235.
14. "Pearl J. Causality: models, reasoning, and inference. 2nd ed. New York: Cambridge University Press, 2009: 350."
15. Rajivlochan M. clinical audits and the state of record keeping in india. Ann Neurosci. 2015 Oct;22(4):197-8. doi: 10.5214/ans.0972.7531.220402. PMID: 26527163; PMCID: PMC4627207.
16. https://www.ema.europa.eu/en/documents/presentation/presentation-regulatory-perspective-real-world-evidence-rwe-scientific-advice-emas-pcwp-hcpwp-joint_en.pdf
17. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en.pdf
18. https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/announcements/optimizing-real-world-evidence-regulatory-decisions.html
19. https://www.getreal-institute.org/
20. ISPE https://pharmacoepi.org/pub/?id=136DECF1-C559-BA4F-92C4-CF6E3ED16BB6
21. ISPOR https://www.ispor.org/strategic-initiatives/real-world-evidence/real-world-evidence-transparency-initiative
22. NICE https://www.nice.org.uk/about/who-we-are/corporate-publications/the-nice-strategy-2021-to-2026
23. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6879703/
24. https://www.futuremedicine.com/doi/10.2217/cer-2020-0228
25. EMA PASS : (Article 1 (15) of Directive 2001/83/EC).https://www.ema.europa.eu/en/human-regulatory/post-authorisation/pharmacovigilance/post-authorisation-safety-studies-pass-0#:~:text=A%20post%2Dauthorisation%20safety%20study%20(PASS)%20is%20defined%20in,product%2C%20or%20of%20measuring%20the

26. PMS : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3148611/

27. RCR: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3853868/

28. Hybrid : https://www.pharmoutsourcing.com/Featured-Articles/180633-Enriched-Real-World-Research-Combining-Existing-and-New-Data-Collection-for-Powerful-Studies/

29. Time frame : https://openheart.bmj.com/content/5/1/e000788

30. Pragmatic trial: https://www.sciencedirect.com/science/article/abs/pii/S0895435609000432

31. External control arm: https://onlinelibrary.wiley.com/doi/full/10.1002/pds.4975

32. Hun-Sung et al. “Real-world Evidence versus Randomized Controlled Trial: Clinical Research Based on Electronic Medical Records.” Journal of Korean medical science vol. 33,34 e213. 26 Jun. 2018, doi:10.3346/jkms.2018.33.e213

33. Food and Drug Administration. Real world Evidence: https://www.fda.gov/media/120060/download

34. Worsley SD, Oude Rengerink K, Irving E, Lejeune S, Mol K, Collier S, Groenwold RHH, Enters-Weijnen C, Egger M, Rhodes T; GetReal Work Package 3. Series: Pragmatic trials and real world evidence: Paper 2. Setting, sites, and investigator selection. J Clin Epidemiol. 2017 Aug;88:14-20. doi: 10.1016/j.jclinepi.2017.05.003. Epub 2017 May 11. PMID: 28502811.

35. Good Clinical Practice Guidance and Pragmatic Clinical Trials: Balancing the Best of Both Worlds Robert J. Mentz, MD1,2, Adrian F. Hernandez, MD, MHS1,2, Lisa G. Berdan, PA, MHS1, Tyrus Rorick, RN1, Emily C. O’Brien, PhD1,2, Jenny C. Ibarra, RN, MSN1, Lesley H. Curtis, PhD1, and Eric D. Peterson, MD, MPH1,2 1Duke Clinical Research Institute, Durham, NC

36. Oude Rengerink K, Kalkman S, Collier S, Ciaglia A, Worsley SD, Lightbourne A, Eckert L, Groenwold RHH, Grobbee DE, Irving EA; Work Package 3 of the GetReal consortium. Series: Pragmatic trials and real world evidence: Paper 3. Patient selection challenges and consequences. J Clin Epidemiol. 2017 Sep;89:173-180. doi: 10.1016/j.jclinepi.2016.12.021. Epub 2017 May 11. PMID: 28502808.

37. Katrien Oude Rengerink, Shona Kalkman, Susan Collier, Antonio Ciaglia, Sally D. Worsley, Alison Lightbourne, Laurent Eckert, Rolf H.H. Groenwold, Diederick E. Grobbee, Elaine A. Irving, Series: Pragmatic trials and real world evidence: Paper 3. Patient selection challenges and consequences, Journal of Clinical Epidemiology, Volume 89,2017,Pages 173-180,ISSN 0895-4356,https://doi.org/10.1016/j.jclinepi.2016.12.021.

38. Thoma A, Farrokhyar F, McKnight L, Bhandari M. Practical tips for surgical research: how to optimize patient recruitment. Can J Surg. 2010;53(3):205-210.

39. https://www.evidera.com/wp-content/uploads/2017/11/Real-World-Evidence-Studies-Need-Patient-Too.pdf

40. Meinecke AK, Welsing P, Kafatos G, Burke D, Trelle S, Kubin M, Nachbaur G, Egger M, Zuidgeest M; work package 3 of the GetReal consortium. Series: Pragmatic trials and real world evidence: Paper 8. Data collection and management. J Clin Epidemiol. 2017 Nov;91:13-22. doi: 10.1016/j.jclinepi.2017.07.003. Epub 2017 Jul 14. PMID: 28716504.

41. Richesson, R. L., & Nadkarni, P. (2011). Data standards for clinical research data collection forms: current status and challenges. Journal of the American Medical Informatics Association : JAMIA, 18(3), 341–346. https://doi.org/10.1136/amiajnl-2011-000107

42. Saczynski JS, McManus DD, Goldberg RJ. Commonly used data-collection approaches in clinical research. Am J Med. 2013;126(11):946-950. doi:10.1016/j.amjmed.2013.04.016

43. Data collection Approach: Commonly Utilized Data Collection Approaches in Clinical Research (nih.gov)

44. ICH Guideline: GUIDELINE FOR GOOD CLINICAL PRACTICE (ich.org)

45. TranCelerate RBM? Introduction - TransCelerate (transceleratebiopharmainc.com)

46. Annas GJ, Grodin MA. The Nazi Doctors and the Nuremberg Code. New York: Oxford University Press; 1992.

47. World Medical Association. Ethical Principles for Medical Research Involving Human Subjects (aka the Declaration of Helsinki). 2004. Available at www.wma.net.

48. Council for International Organizations of Medical Sciences (CIOMS). International ethical guidelines for biomedical research involving human subjects. Geneva, Switzerland: CIOMS; 2012.

49. World Medical Association (WMA). WMA Declaration of Helsinki: ethical principles for medical research involving human subjects. JA-MA 2013; 310:2191-4.

50. Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations, and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use, (2001).

51. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC (2014).

52. Weijer C, Grimshaw JM, Eccles MP, McRae AD, White A, Brehaut JC, et al. The Ottawa statement on the ethical design and conduct of cluster randomized trials. PLoS Med 2012;9: el001346.

53. Turner A., Ethical issues raised by Big Data and Real World Evidence projects Oxford Internet Institute, University of Oxford Dec 2017

54. US FDA Use of Electronic Informed Consent Questions and Answers Guidance for Institutional Review Boards, Investigators, and Sponsors Guidance document, Dec 2016

55. New Drugs and Clinical Trial Rules 2019 under the Indian “Drug and Cosmetics Act 1940”

56. Kalkman S. et al, Informed Consent for Randomized Control Trial challenges and alternative informed consent proposal for RWE J. Clin. Epidemiology, 89 (2017) 181-187.

57. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA 2003;290: I 624-32.

58. Sugarman J, Califf RM. Ethics and regulatory complexities for pragmatic clinical trials. JAMA 20 I 4;3 I :2381-2.

59. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147.

60. Eichler HG, Abadie E, Breckenridge A, Flamion B, Gustafsson LL, Leufkens H, et al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. Nat Rev Drug Discov 2011; 10:495-506.

61. SJ Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Clin Epidemiol 2009; 62:499-505.

62. Califf RM, Sugarman J. Exploring the ethical and regulatory issues in pragmatic clinical trials. Clin Trials 20 IS; 12:436-41.

63. Drugs and Cosmetics Act. Available at: Drugs_CosmeticsAct1940_Rules1945.pdf (cdsco.gov.in) [Last accessed on 10th Oct 2021]

64. National Ethical Guidelines for Biomedical and Health Research involving Human participants available at: ICMR_National_Ethical_Guidelines.pdf (ncdirindia.org) [Last accessed – 10th Oct 2021]

65. Jie Chen, Martin Ho, Kwan Lee et al. The Current Landscape in Biostatistics of Real-World Data and Evidence: Clinical Study Design and Analysis, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2021.1883474

66. Hiemstra, B., Keus, F., Wetterslev, J. et al. DEBATE-statistical analysis plans for observational studies. BMC Med Res Methodol 19, 233 (2019). https://doi.org/10.1186/s12874-019-0879-5

67. McMurry TL, Hu Y, Blackstone EH, Kozower BD. Propensity scores: Methods, considerations, and applications in the Journal of Thoracic and Cardiovascular Surgery. Thorac Cardiovasc Surg. 2015 Jul; 150(1):14-9.

68. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. STROBE initiative: strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335(7624):806–8.

69. Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med. 2013;35(2):121-126. doi:10.4103/0253-7176.11623

70. Organized structure of real-world evidence best practices: moving from fragmented recommendations to comprehensive guidance. Ashley Jaksa, James Wu, Páll Jónsson, Hans-Georg Eichler, Sarah Vititoe, and Nicolle M Gatto. Journal of Comparative Effectiveness Research 2021 10:9, 711-731

71. Enhancing the QUAlity and Transparency Of health Research. September 2021. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Accessed on 27th Oct 2021. https://www.equator-network.org/reporting-guidelines/strobe/

72. GRACE Principles: Recognizing High-Quality Observational Studies of Comparative Effectiveness. June 4, 2010. Nancy A. Dreyer, PhD, Sebastian Schneeweiss, MD, Barbara J. McNeil, MD, Marc L. Berger, MD, Alec M. Walker, MD, Daniel A. Ollendorf, MPH, Richard E. Gliklich, MD, for the GRACE Initiative. The American Journal of Managed Care, June 2010, Volume 16, Issue 6

73. U. S. Food and Drug Administration. May 2013. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets. Accessed on 26th Oct 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/best-practices-conducting-and-reporting-pharmacoepidemiologic-safety-studies-using-electronic.

74. Velentgas P, Dreyer NA, Nourjah P, et al., editors. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Jan. Available from: https://www.ncbi.nlm.nih.gov/books/NBK126190/?report=classic

75. International Society for Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP) (2015). https://www.pharmacoepi.org/resources/policies/guidelines-08027/Accessed on 25th Oct 2021.

76. European Medicines Agency. Guidance for the format and content of the protocol of non-interventional post-authorisation safety studies (2012). Accessed on 24th Oct 2021. https://www.ema.europa.eu/en/documents/other/guidance-format-content-protocol-non-interventional-post-authorisation-safety-studies_en.pdf

77. National Digital Health Mission: Health Data Management Policy. Accessed at https://ndhm.gov.in/health_management_policy on 23 Sep 2021.

78. Saksena N, Matthan R, Bhan A, et al, Rebooting consent in the digital age: a governance framework for health data exchangeBMJ Global Health 2021;6:e005057.

79. Data Protection. Data intelligence. Journal of Intellectual Property Law & Practice, 2011, Vol. 6, No. 12. Pp. 863-856.

80. Pai, M.M.M., Ganiga, R., Pai, R.M. et al. Standard electronic health record (EHR) framework for Indian healthcare system. Health Serv Outcomes Res Method 21, 339–362 (2021).

81. Implementation of Cloud based Electronic Health Record (EHR) for Indian Healthcare Needs. R. Kavitha*, E. Kannan and S. Kotteswaran. Indian Journal of Science and Technology, Vol 9(3), DOI: 10.17485/ijst/2016/v9i3/86391, January 2016

82. Sen, Poulomi, EU GDPR and Indian Data Protection Bill: A Comparative Study (April 26, 2021). Available at SSRN: https://ssrn.com/abstract=3834112 or http://dx.doi.org/10.2139/ssrn.3834112 accessed on 20 Sep 2021

83. Misra, Rajat and Grover, Rajat, Future of Privacy: Evaluating the Personal Data Protection Bill, 2019 in Light of Contract for the Web (April 8, 2020). Available at SSRN: https://ssrn.com/abstract=3592363 or http://dx.doi.org/10.2139/ssrn.3592363. Accessed on 20 sep 2021

84. Wang SV, Pinheiro S, Hua W, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. BMJ. 2021;372:m4856. Published 2021 Jan 12. doi:10.1136/bmj.m4856.

85. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10):e1001885. Published 2015 Oct 6. doi:10.1371/journal.pmed.1001885

86. GRACE Initiative. GRACE Principles. 2010. Available at: https://www.graceprinciples.org/doc/GRACE_Principles.pdf. Accessed on 16 Nov 2021.

87. Public Policy Committee, International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practice (GPP). Pharmacoepidemiol Drug Saf. 2016;25(1):2-10. doi:10.1002/pds.3891.

88. Berger ML, Martin BC, Husereau D, et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report [published correction appears in Value Health. 2014 Jun;17(4):489]. Value Health. 2014;17(2):143-156. doi:10.1016/j.jval.2013.12.011/

89. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007;4(10):e297. doi:10.1371/journal.pmed.0040297.