VALUE IN HEALTH ▮ (2017) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

What Is Real-World Data (RWD)? A Review of Definitions Based on Literature and Stakeholder Interviews

Amr Makady, MSc[1,2,], (on behalf of GetReal Work Package 1), Anthonius de Boer, MD, PhD[2], Hans Hillege, PhD[3], Olaf Klungel, PhD[2], Wim Goettsch, PhD[1,2]*

[1]*The National Healthcare Institute (ZIN), Diemen, The Netherlands;* [2]*Department of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht, The Netherlands;* [3]*Department of Epidemiology, University Medical Centre Groningen, Groningen, The Netherlands*

ABSTRACT

Background: Despite increasing recognition of the value of real-world data (RWD), consensus on the definition of RWD is lacking. **Objectives:** To review definitions publicly available for RWD to shed light on similarities and differences between them. **Methods:** A literature review and stakeholder interviews were used to compile data from eight groups of stakeholders. Data from documents and interviews were subjected to coding analysis. Definitions identified were classified into four categories: 1) data collected in a non-randomized controlled trial setting, 2) data collected in a non-interventional/non-controlled setting, 3) data collected in a non-experimental setting, and 4) others (i.e., data that do not fit into the other three categories). The frequency of definitions identified per category was recorded. **Results:** Fifty-three documents and 20 interviews were assessed. Thirty-eight definitions were identified: 20 out of 38 definitions (53%) were category 1 definitions, 9 (24%) were category 2 definitions, 5 (13%) were category 3 definitions, and 4 (11%) were category 4 definitions. Differences were identified between, and within, definition categories. For example, opinions differed on the aspects of intervention with which non-interventional/non-controlled settings should abide. No definitions were provided in two interviews or identified in 33 documents. **Conclusions:** Most of the definitions defined RWD as data collected in a non-randomized controlled trial setting. A considerable number of definitions, however, diverged from this concept. Moreover, a significant number of authors and stakeholders did not have an official, institutional definition for RWD. Persisting variability in stakeholder definitions of RWD may lead to disparities among different stakeholders when discussing RWD use in decision making.

Keywords: definitions, real-world data, real-world evidence, real-world studies, review, stakeholder definitions.

Introduction

Randomized controlled trials (RCTs) provide the ideal study design for demonstrating causality between the use of a specific medicine and intended and unintended effects under ideal conditions. In conventional RCTs conducted during phase III drug development, patients are based on stringent inclusion and exclusion criteria and subsequently randomized to different treatment arms to counteract the influence for known and unknown confounders [1,2]. In addition, monitoring and follow-up procedures for trial subjects are often highly controlled [1,2].

The highly selective populations examined within the setting of RCTs are often not comparable with the more heterogeneous populations in clinical practice in which medicines are administered to patients with varying genetic make-ups, who present with different comorbidities or already receive different medications for other morbidities. Consequently, experimental medicines being presented for marketing authorization are accompanied by data that provide efficacy as well as safety data with very high internal validity but whose results may not be easily generalizable to a broader, more heterogeneous population [2]. This disparity of findings on the therapeutic efficacy of medicines from tightly controlled RCT settings and the effectiveness of medicines in the real world has been previously defined by Eichler et al. [3] as the "efficacy-effectiveness gap."

Regulatory agencies are thus faced with the issue of making decisions on the basis of data with inherent uncertainties on the aspects of real-world effectiveness. Similarly, health technology assessment (HTA) agencies and health care payers conventionally exploit RCT-generated evidence available at the time of initial reimbursement decisions to assess the relative effectiveness of new products. As a result, many stakeholders such as the pharmaceutical industry, regulatory agencies, HTA agencies, and payers have begun exploring options for the use of real-world data (RWD) as a complementary source to RCT data for establishing a more robust evidence base on the effectiveness of medicines, as well as the relative effectiveness compared with existing products in clinical practice [4,5].

* *Address correspondence to:* Amr Makady, The National Healthcare Institute (ZIN), Eekholt 4, Diemen 1112 XH, The Netherlands.
E-mail: amakady@zinl.nl.

1098-3015$36.00 – see front matter Copyright © 2017, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jval.2017.03.008

Table 1 – ISPOR, ABPI, RAND Corporation, and IMI-GetReal definitions for RWD.

Term and source	Definition
RWD (ISPOR [7])	Data used for decision making that are not collected in conventional RCTs.
RWD (ABPI [8])	For the purposes of this guidance, "RWD" will refer to data obtained by any non-interventional methodology that describe what is happening in normal clinical practice.
RWD (RAND [9])	"RWD" is an umbrella term for different types of health care data that are not collected in conventional RCTs. RWD in the health care sector come from various sources and include patient data, data from clinicians, hospital data, data from payers, and social data.
RWD (IMI-GetReal [10])	An umbrella term for data regarding the effects of health interventions (e.g., benefit, risk, and resource use) that are not collected in the context of conventional RCTs. Instead, RWD are collected both prospectively and retrospectively from observations of routine clinical practice. Data collected include, but are not limited to, clinical and economic outcomes, patient-reported outcomes, and health-related quality of life. RWD can be obtained from many sources including patient registries, electronic medical records, and observational studies.

ABPI, Association of the British Pharmaceutical Industry; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; RCT, randomized controlled trial; RWD, real-world data.

In addition, RWD are currently used during drug development to examine aspects such as the natural history of a disease, delineating treatment pathways in clinical practice, determining the costs and resource use associated with treatment interventions, and determining outcomes related to comparator interventions [4,6]. Such knowledge may inform aspects of early drug development such as clinical trial design or the comparative effectiveness of comparator treatments within a given indication.

Despite the increasing popularity of RWD collection and use for drug development, drug regulation, and HTA, a certain degree of disparity remains among different stakeholders when it comes to thoroughly defining RWD [6]. Therefore, this study aimed to conduct a review of definitions for RWD available in literature and stakeholders' definitions of the term within the context of drug development, drug regulation, and HTA of pharmaceutical products to straighten out the similarities and differences between them. In addition, the article will review which data sources stakeholders believe as being RWD and which study designs they consider to generate RWD. Subsequently, the article will shed light on existing definitions for the term RWD developed by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) [7], the Association of the British Pharmaceutical Industry (ABPI) [8], RAND Corporation [9], and the IMI-GetReal consortium [10] (see Table 1).

Methods

Two qualitative methods were used to compile data from relevant stakeholders: a literature review and stakeholder interviews. Data compilation from eight stakeholder groups was performed, namely, HTA agencies, the pharmaceutical industry, regulatory agencies, academia, health care providers, health care insurers/payers, patient organizations, and initiatives using, or commissioning research on, RWD (e.g., ISPOR and the Patient-Centered Outcomes Research Institute).

For the literature review, PubMed was used to search scientific literature from January 1, 2005, to December 31, 2016 (date of search). The search strategy used is presented in Appendix Figure i in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.03.008. To locate gray literature, Web sites belonging to the eight stakeholder groups were consulted (see Appendix Table i in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.03.008 for a list of Web sites consulted). Search functions on stakeholder Web sites were used when available, using terms such as "real-world data," "real-world evidence," "clinical effectiveness data," "real-world outcome," "comparative effectiveness," or "relative effectiveness." Search results from both scientific and gray literature were independently screened by two of the authors according to predefined inclusion and exclusion criteria (see Appendix Table ii in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.03.008). Any discrepancies for inclusion and exclusion of articles were resolved by consensus between the two authors.

A standardized data abstraction form was created in Microsoft Excel and used to locate information in the documents selected after screening. Data elements included in the data abstraction form were author name(s), publication year, the type of document, definition(s) of RWD provided, and data sources considered as RWD and study designs considered to generate RWD (e.g., claims databases and observational studies, respectively). Two of the authors extracted data independently from the selected documents. Any discrepancies in the extracted data were resolved by consensus between the two authors.

With regard to stakeholder interviews, stakeholders from the eight previously mentioned groups were selectively sampled on the basis of seniority and function, with a preference for senior representatives involved in work on RWD use within their respective organizations. Information for identifying representatives was retrieved from stakeholder Web sites and/or the authors' professional network. All representatives were approached by email using a standardized invitation to participate in semistructured interviews. To increase the validity of stakeholder views, participants were provided the freedom to invite colleagues they deemed relevant to take part in the interviews. Tailored questionnaires were developed for each stakeholder group and sent to stakeholders who agreed to participate 2 weeks before the interview to guide discussions (see Appendix Figures ii to iv in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.03.008 for examples of questionnaires sent to three stakeholder groups). Interviews were conducted, recorded, and subsequently transcribed for further analysis.

The sampling of stakeholders and interview protocols were compared with recommendations in the consolidated criteria for reporting qualitative studies (COREQ) [11] to ensure good quality. The COREQ checklist provides guidance for explicit and comprehensive reporting of qualitative studies using interviews and focus groups.

It is important to note that the interviews were conducted as part of a larger study on policies and perspectives on RWD [6], and thus the scope of questions posed during the interviews extended beyond the definition of RWD. All questionnaires, however, included the following three questions: