VALUE IN HEALTH 19 (2016) 75–81

Available online at www.sciencedirect.com

ScienceDirect

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

Comparative Effectiveness Research/Health Technology Assessment (HTA)

The "Efficacy-Effectiveness Gap": Historical Background and Current Conceptualization

Clementine Nordon, MD PhD[1,], Helene Karcher, PhD[2], Rolf H.H. Groenwold, PhD[3], Mikkel Zöllner Ankarfeldt, PhD[4], Franz Pichler, PhD[5], Helene Chevrou-Severac, PhD[6], Michel Rossignol, MD, PhD[7], Adeline Abbe, MSc[8], Lucien Abenhaim, MD, PhD[2], on behalf of the GetReal consortium*

[1]*LASER Research, Paris, France;* [2]*LASER Analytica, London, UK;* [3]*Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, The Netherlands;* [4]*Novo Nordisk A/S, Soborg, Denmark;* [5]*Eli Lilly and Company, Melrose Park, Australia;* [6]*Takeda Pharmaceuticals International, Glattpark-Opfikon, Switzerland;* [7]*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada;* [8]*Sanofi R&D, Chilly Mazarin, France*

ABSTRACT

Background: The concept of the "efficacy-effectiveness gap" (EEG) has started to challenge confidence in decisions made for drugs when based on randomized controlled trials alone. Launched by the Innovative Medicines Initiative, the GetReal project aims to improve understanding of how to reconcile evidence to support efficacy and effectiveness and at proposing operational solutions. **Objectives:** The objectives of the present narrative review were 1) to understand the historical background in which the concept of the EEG has emerged and 2) to describe the conceptualization of EEG. **Methods:** A focused literature review was conducted across the gray literature and articles published in English reporting insights on the EEG concept. The identification of different "paradigms" was performed by simple inductive analysis of the documents' content. **Results:** The literature on the EEG falls into three major paradigms, in which EEG is related to 1) real-life characteristics of the health care system; 2) the method used to measure the drug's effect; and 3) a complex interaction between the drug's biological effect and contextual factors. **Conclusions:** The third paradigm provides an opportunity to look beyond any dichotomy between "standardized" *versus* "real-life" characteristics of the health care system and study designs. Namely, future research will determine whether the identification of these contextual factors can help to best design randomized controlled trials that provide better estimates of drugs' effectiveness.

Keywords: efficacy-effectiveness gap, pragmatic clinical trials, outcomes research, pharmaceuticals.

Introduction

Regulatory approval of a new drug requires evidence of a positive efficacy-safety ratio: the extent to which the drug does more good than harm [1] is usually measured using randomized controlled trials (RCTs). Prelaunch drug development relies heavily on RCTs. When new drugs are launched in the market, little is known about their impact under routine prescribing practice and utilization of drugs, conventionally known as effectiveness [2]. The concept of the "efficacy-effectiveness gap" (EEG) describes possible discrepancies and complementary scientific evidence on efficacy and effectiveness. The awareness raised around this concept [3–5] results from how it may impact clinical and policy decisions on drugs. Research initiatives that aim to improve understanding of how evidence of efficacy and effectiveness can be reconciled and introduced at an earlier stage of drug development have been launched worldwide [6,7].

In the context of the GetReal project [7], the objectives of this narrative review on the EEG were 1) to understand the historical background in which the concept of EEG has emerged and 2) to describe the conceptualization of EEG.

Methods

A narrative-focused literature review of documents published in English to synthesize knowledge on EEG and address the objectives set out by the authors was carried out.

Identification of Documents

First, a broad search of the gray literature was performed across Internet Web sites from governmental authorities [8,9] and nongovernmental initiatives for EEG [6,9–12], as well as

* *Address correspondence to:* Clementine Nordon, LASER Research, 10, place de Catalogne, 75014 Paris, France.
E-mail: clementine.nordon@la-ser.com.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2015.09.2938

Table 1 – Search algorithm on the "efficacy-effectiveness gap".

Efficacy		
1	Efficacy	efficacy [Title/Abstract]
2	Clinical trials	"Clinical Trial, Phase II" [Publication Type] OR "Clinical Trial, Phase III" [Publication Type] OR "Clinical Trial, Phase IV" [Publication Type] OR "Controlled Clinical Trial" [Publication Type] OR "Multicenter Study" [Publication Type] OR "Randomized Controlled Trial" [Publication Type] OR "Clinical Trials, Phase II as Topic" [Mesh] OR "Clinical Trials, Phase III as Topic "[Mesh] OR "Clinical Trials, Phase IV as Topic" [Mesh] OR "Controlled Clinical Trials as Topic" [Mesh] OR "Multicenter Studies as Topic" [Mesh] OR clinical trial*[Title/Abstract] OR clinical trial*[Text Word] OR "randomized" [Title/Abstract] OR "randomized" [Title/Abstract] OR "randomized" [Text Word] OR "randomised" [Text Word] OR "randomization" [Title/Abstract] OR "randomization" [Title/Abstract] OR "randomization" [Text Word] OR "randomization" [Text Word] OR "random allocation" [Title/Abstract] OR "double blind" [Title/Abstract] OR "double blinded" [Title/Abstract] OR "double masked" [Title/Abstract] OR "single blind" [Title/Abstract] OR "single blinded" [Title/Abstract] OR "single masked" [Title/Abstract]
3	Explanatory trials	Explanatory [Title/Abstract] AND (trials [Title/Abstract] OR trial [Title/Abstract] OR studies [Title/Abstract] OR study [Title/Abstract])
4		1 AND (2 OR 3)
Effectiveness		
5	Effectiveness	Effectiveness [Title/Abstract] OR "translational research" [Title/Abstract]
6	Real life	("Real Life" [Title/Abstract]) AND (study [Title/Abstract] OR studies [Title/Abstract])) OR ("real word" AND study) [Title/Abstract] OR ("real word" AND data*) [Title/Abstract] OR (observational [Title/Abstract] AND (study [Title/Abstract] OR studies [Title/Abstract])) OR ("Real Life" [Title/Abstract] AND conditions [Title/Abstract]) OR naturalistic [Title/Abstract] OR "patient-oriented research" [Title/Abstract]
7	CER	"comparative effectiveness" [Title/Abstract] OR CER [Title/Abstract]
8	Pragmatic studies	(Pragmatic [Title/Abstract] OR Practical [Title/Abstract]) AND (trials [Title/Abstract] OR trial [Title/Abstract] OR studies [Title/Abstract] OR study [Title/Abstract])
9		5 AND (6 OR 7 OR 8)
Concept		
10	Gap	Gap*[Title/Abstract] OR issue*[Title/Abstract] OR complexit*[Title/Abstract] OR barrier*[Title/Abstract] OR facilitator*[Title/Abstract] OR problem*[Title/Abstract] OR bias [Title/Abstract]
11		4 AND 9 AND 10

CER, cost-effectiveness research.

Google and Google Scholar (for books and theses). This approach identified key authors and the main keywords to be used subsequently in a focused literature search (second step). This search was performed applying an algorithm (see Table 1) to the Embase, PubMed, and Cochrane Collaboration Web sites. Finally, the search was completed using a "snowball method" across the reference lists from the articles identified. The search ended when saturation was reached and no additional information relating to the identified paradigms could be extracted.

Selection and Appraisal of Documents

Documents were selected when 1) providing key elements about the historical background in which the EEG concept emerged and 2) when reporting insights (definitions and solutions) on the EEG concept. Quality appraisal of the document was performed ensuring "relevance" and "clarity" of the research questions, both regarding the qualitative research [13] and unprompted judgment, that is, our expertise [14].

Analysis and Data Synthesis

To identify different "paradigms," all the selected documents were used to extract and synthetize how the phenomenon of interest (EEG) was understood or discussed (the authors' perspective). A simple inductive analysis of the documents' content was performed because no paradigm classification had been defined before our study. The identification of different items took place as the data were being collected and analyzed. One researcher (C. N.) read the documents and annotated the main items in the margins. Subsequently, other researchers were involved (L.A. and M.R.), proceeding to a more analytical listing of items and the identification of paradigms.

Results

Search Results

The search algorithm allowed the identification of 672 academic manuscripts after removing duplicates. These were screened on title, and for a further 100 articles, the abstracts or the full manuscript was read.

Historical Background

The growing number of available medical technologies since the 1950s and the increasing amount of related information have generated a need for rigorous assessment of the potential benefit [15], information synthesis, and knowledge dissemination [16] of medical interventions.

Standardization of Trials and Information Synthesis

The standardization of clinical trials started in the 1950s with the implementation of founding principles for RCTs [17], but the importance of basing the assessment of therapies through clinical trials—rather than relying on prescribers' "opinions" [18]—reached the wider medical community nearly 20 years later. During this period, the concept of "attitude" [19] in a trial (explanatory *vs.* pragmatic) was introduced: the "explanatory attitude" related to the objective of acquiring information about the efficacy of drugs, whereas the "pragmatic attitude" related to the objective of gaining information about their effectiveness. These two attitudes were described as leading to two ways of interpreting results: understanding *versus* making decisions. The

"explanatory" attitude alone, however, has prevailed for more than 20 years.

In turn, methods for synthetizing the information from several clinical trials (meta-analyses) were standardized between 1980 and the early 1990s [20]. The Cochrane Collaboration [21] was created to facilitate the development of systematic reviews and meta-analyses of RCTs.

Knowledge dissemination

The creation of the American Consensus Development Program [22] in 1977 was an early attempt to improve the dissemination of biomedical research into clinical practice. Similarly, the paradigm of evidence-based medicine (EBM) emerged to promote good clinical practice [23] and was presented as "a practice integrating individual clinical expertise with the best available external clinical evidence from systematic research" [24]. (Fig. 1).

The EEG

Notwithstanding the standardization of clinical trials, methods to synthetize evidence, and platforms for knowledge dissemination, a chasm between research and clinical practice [25] has been noted and called the "knowledge gap" [22] or the "efficacy-effectiveness gap" [26]. Lehman et al. [26] explained: "Effectiveness refers to the impact of treatment under usual treatment conditions, in which patient factors […], provider factors […], and service system factors […] that can affect treatment outcomes, are not controlled."

Over the years, several ways of understanding the EEG have been suggested. Through our literature review, we have identified that the EEG concept falls into three major paradigms (see Table 2) relating to the role of 1) health care settings, 2) methods used to assess drugs' effect, and 3) the interaction between the drug and contextual factors.

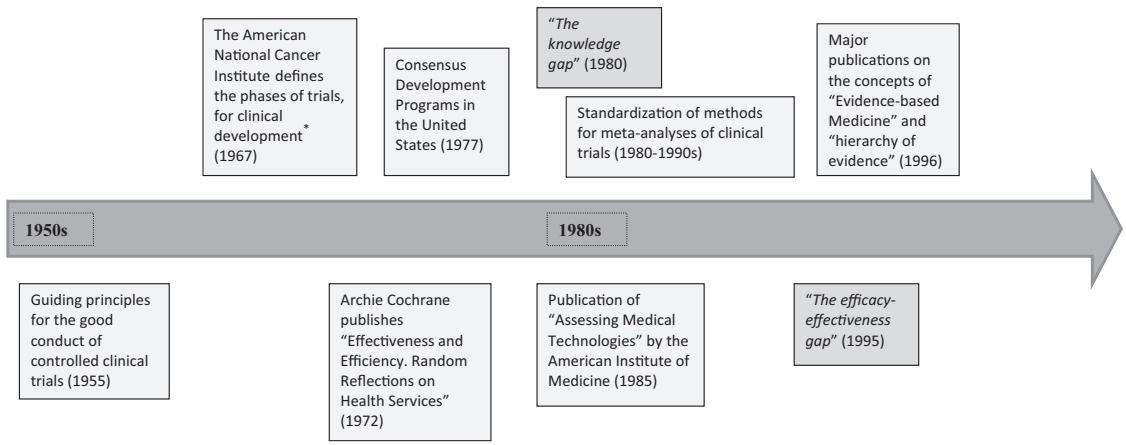
First Paradigm: The Issue of Behavior Toward Drugs' Prescription and Use

In a first paradigm, the EEG may be explained by characteristics of health care settings, including the "behavior" of caregivers and patients' adherence to treatment. In this paradigm, the optimal effect of the drug is thought to be distorted by real-life factors; namely, the EEG is understood as an issue of *behavior*.

First, the physicians' behavior has been thought to jeopardize the effectiveness of drugs and the dissemination of knowledge. One example can be found in a book published by the American Institute of Medicine, in which the authors stated that "patterns of medical practice often diverge from recommendations based on controlled clinical evaluations" [27, p. 176] and explained that one reason for this relates to physicians' behavior, defined as a combination of skepticism about innovation, lack of medical training, characteristics of the practice setting, and so forth. However, RCTs were thought to be the strongest method of assessing the efficacy of drugs. Similarly, it has been suggested that cognitive biases in interpreting new results (e.g., the tendency to embrace evidence supporting preconceived ideas) may compromise the generation and dissemination of evidence on efficacy into practice [28]. It is noteworthy that because clinical research and medical care involve human beings—physicians and patients—subjectivity cannot be set aside [29].

Second, the patients' adherence (i.e., the extent to which a patient's behavior corresponds to agreed recommendations from a health care provider [30]) has been suggested as a key factor to explain EEG, especially in chronic disorders [31–33]. The underlying rationale is that in clinical trials, attempts are made to maximize adherence, which would not be the case in daily practice. Numerous adherence-enhancement strategies have been developed (cognitive-educational interventions, self-reporting devices, or even telephone coaching [34]). Their impact on medication adherence, however, was found to be modest and not sustained over time [35]. In turn, the impact of enhancing medication adherence on effectiveness is yet to be established. Also, the relationship between the effect of a drug and drug adherence is complex and may apply in both directions: bad adherence by patients may lower the impact of treatments, and the poor efficacy (or tolerance) of a drug as perceived by the patient may lower the patient's adherence, in particular for medications that are expected to have a direct and identifiable effect on symptoms. In recognition of this intricate relationship, the investigators of a large pragmatic trial on antipsychotic drugs [36] have used the "discontinuation of treatment for any cause" as the primary outcome, explaining that "this measure integrates patients' and clinicians' judgments of efficacy, safety, and tolerability into a global measure of effectiveness […]."

Finally, the disparity in resource and access to care between settings has been considered as being responsible for the EEG [37,38]. For example, in elderly people with depression [39], the EEG has been described as the result of a combination of health care–related factors and intricate barriers, categorized into "patient barriers" (e.g., high costs of care, misdiagnosis of



* Zubrod, C.G. Multiclinic trials in cancer chemotherapy. *Can Med Assoc J* **97**, 101-3 (1967)

Fig. 1 – Historical background of the efficacy-effectiveness gap concept.

Table 2 – Summary of the three paradigms identified.

Paradigm	Description	Themes encompassed
1. The EEG is related to real-life characteristics of the health care system	The ideal effect of the drug is distorted by real-life characteristics of the health care system, related to the physician, the patient, and access to health care resources	• In routine practice, the *physicians'* "behavior" regarding medical guidelines and dissemination of knowledge is not optimal • In routine practice, the *patients' adherence* is not optimal • In routine practice, there are access barriers to health care resources
2. The EEG is related to an issue of the method used to measure the drug's effect	Efficacy and effectiveness studies use different study designs and design parameters, hence the EEG	• Concept of *evidence-based medicine* and *hierarchy of evidence*: the efficacy is the *real* effect of the drug; the RCTs are the criterion standard for measuring the drug's effect • Concept of *pragmatism*: RCTs' lack of generalizability; any direct dissemination of evidence coming from clinical trial into clinical practice is inadequate
3. The EEG is related to an issue of complex interaction	The drug's effect is the result of complex (and multiple) interactions between the biological effect of the drug and "real-life" contextual factors, hence the EEG	• Some contextual factors are (significantly) interacting with the drug's biological effect ("drivers of effectiveness") • An imbalance in the distribution of these factors between efficacy and effectiveness studies may cause an EEG

EEG, efficacy-effectiveness gap; RCT, randomized controlled trial.

depression in elderly people, and older people rarely seeking care), "provider barriers" (e.g., competing care of comorbidities), and "policy barriers" (e.g., geographic distance from care provider). It is believed that facilitating access to care could improve medical outcomes, as exemplified in a study on the "easy access to ambulatory service" program by patients with bipolar disorder [40].

Second Paradigm: The Issue of Method Used and Measure

A second paradigm holds that the EEG may be explained by the different study designs and methods being used to assess the impact of drugs (efficacy vs. effectiveness studies). In short, the EEG is understood here as an issue of *method/measure* in which the difference in effect size would stand in a difference relating to the study design.

There is indeed an association between the choice in study design parameters and the drug's effect size, as evidenced by Naudet et al. [41], who have investigated the impact of several design parameters of RCTs *versus* observational studies on the effect size of antidepressants. One of the reported results referred to the use of double blinding (*vs.* none), the comparison against placebo (*vs.* active comparator), or the use of intention-to-treat analyses (*vs.* per-protocol) as being associated with a lower effect size, regardless of patients and disease characteristics.

One finding of our literature review, however, is that within this paradigm, two concepts are opposed [42]: the concept of "hierarchy of evidence" supporting the superiority of RCT (due to a supposedly better internal validity) over nonrandomized studies, and the concept of "pragmatism" arguing that RCTs would lack generalizability (external validity) as compared with more pragmatic studies.

The Concept of Hierarchy of Evidence

The development of EBM and the concept of hierarchy of evidence might have introduced ambiguity regarding the role of RCTs within the frame of health technology assessment, with the underlying assumption that efficacy would be the *real* effect of the drug whereas effectiveness would be a *distorted* one, derived from "biased" real-life observational studies [43,44]. In a publication on EBM [24], its authors simultaneously stated that RCT "has become the gold standard for *judging* whether a treatment does more good than harm" and that EBM "is the conscientious, explicit, and judicious use of current best evidence in *making decisions* about the care of individual patients," which might have blurred the essential difference between assessing drugs' efficacy and choosing between drugs in clinical practice [19]. Furthermore, the concept of hierarchy of evidence [45] has consolidated RCTs and meta-analyses of RCTs as the leaders among type of studies aiming at assessing drugs' impact. This concept holds that the quality of study designs is set in stone, irrespective of the study objective, that is, whether it is intending to measure the biologic effect of the drug or not, or informing prescribers or policymakers on the best options to choose from in routine practice. Although a more balanced use of this hierarchy of evidence has been recently suggested [46,47], the concept of hierarchy of evidence may have contributed to the idea that "real-life" studies are not as legitimate as RCTs to provide evidence on the effect of drugs.

The Concept of "Pragmatism"

Conversely, it has been argued that RCTs lack generalizability. The "cautious generalization" of clinical trials was emphasized a long time ago [17], as well as the necessity to interpret results of clinical trials within the context of routine practice, using, for instance, the physicians' expertise [22]. The limited generalizability of RCTs, however, has been explained in more detail only recently. Many aspects of the traditional RCT design do not represent routine clinical practice: the selection of the population [48–50], its homogeneity [5], interventions not reflecting routine practice [51], and outcomes often not relevant enough for clinicians and patients [25]. The concept of pragmatism holds that this lack of generalizability has led to an EEG and that any direct dissemination of evidence arising from clinical trials into clinical practice might be inadequate [52,53]. In line with this concept, the generation of real-life evidence on the impact of drugs is becoming increasingly recommended, namely, for

pharmaceutical companies [10,51,54], in which such evidence is seen as complementary [55–57].

Third Paradigm: The Issue of Increased Variability and Interaction

More recently, a third paradigm has been developed in which any quantitative difference in drugs effect estimates, as measured in an experimental setting or in routine practice, may be understood as the result of interaction of multiple real-life characteristics on the purely biological effect of the drug. Eichler et al. [5] have explained that "to a large extent, the EEG may be considered a result of increasing variability of drug response owing to a combination of genetic, other biological and behavioral factors." The factors of potentially increasing variability in real life were categorized into 1) intrinsic biological characteristics of patients (genetics, physiology, comorbidities, etc.); 2) extrinsic environmental factors (diet, air pollution, health care system characteristics, etc.); and 3) behavioral factors (off-label prescriptions, patient adherence, etc.).

This paradigm encompasses the first two paradigms described earlier and also brings the EEG concept into a more operational level: if the EEG is related to the increasing variability (or the modification) of real-life factors, which statistically speaking corresponds to effect modification and/or interaction, then this gap may not only be explained but can also be anticipated and predicted.

To illustrate this paradigm, we report one study by Schneeweiss et al. [58], who have applied the patients' eligibility criteria commonly used in RCTs on statins to an observational study population, by sequentially excluding prevalent drug users, patients with contraindications, or patients with low adherence and so forth. They have examined the extent to which the 1-year mortality rate ratio estimates were changed and evidenced that using more restrictive eligibility criteria modified the mortality rate ratio estimates to the levels found in RCTs. These results suggest the existence of some interaction effect of patients' characteristics on the association between statins and mortality. On the same line, Ankarfeldt et al. [59] explored the impact of a high protein diet on weight change when RCTs and observational studies showed conflicting results. The authors suggested that being overweight and obese—which characterize the patients included in RCTs but not to the same extent, the patients included in observational studies—can act as an effect modifier on the association between a high protein diet and weight change. Another illustration is provided by Chassang et al. [60] who explored the impact of blinding (vs. open label) on the effect size of antidepressants. In line with Naudet et al. [41], the authors found that in double-blind RCTs, the treatment response tends to be smaller than in open-label studies because there is no "patients' beliefs effect" (expectancies over active treatment, leading to a patients' change in behavior) in double-blind RCTs. Although not new, the study results demonstrate—both through mathematical formalization and through empirical data analyses—that a difference in treatment effect size can happen only in the presence of interaction. This article provides a clear and formal demonstration of the role of interaction in the EEG.

Discussion

The present article is the result of extensive literature review on the EEG focusing on the current understanding of this concept.

Main Results

The present narrative study is thought to be the first to review the literature on the EEG, so as to reflect the whole spectrum of its conceptualization and to elicit the scientific paradigms that have influenced physicians and decision makers for decades.

Three Main Paradigms Render the EEG

A first paradigm holds that imperfections of the health care system in real life may explain why effectiveness outcomes are disappointing compared with efficacy outcomes. Thus, it is believed that all aspects of care in real life should be brought up to the standards of the experimental setting, with several levels of intervention: medical guidelines, knowledge dissemination, adherence-enhancement strategies, and so forth. These interventions are certainly commendable efforts to be made. The actual impact of such interventions, however, would be difficult to assess. Of note, this paradigm is not considering the possibility of the effectiveness being superior to the efficacy. One can imagine that the patient-physician relationship is a lever to patients' adherence, hence a better outcome in real life. Similarly, it is possible (and desirable) that in routine practice, physicians choose between therapeutic options so as to maximize the chance of success in one particular patient. A second paradigm advocates that the methodology used to measure a drug's effect (efficacy vs. effectiveness) has an impact on the drug's outcome, hence the EEG. Although this is probably often the case, one output of this review is that most of the debate has been focused on the pros and cons of each study design option. This debate is somewhat sterile because efficacy and effectiveness studies actually address different but complementary questions. In their pursuit of recognizing the complementary importance of explanatory and pragmatic designs, some authors have introduced the idea of "the explanatory-pragmatic continuum" [61]. Finally, a third paradigm has been identified, championing that a drug's therapeutic impact is the result of complex interactions between the drug's biologic effects and contextual factors (patient-related, provider-related, and health care–related). These interactions are at play whatever the study setting or design, but may differ in magnitude as a result of a drug being assessed in standardized (RCT) or in more flexible (real-life) conditions, hence the discrepancies in results. This third paradigm encompasses the first two and provides the opportunity to look beyond any dichotomy between standardized *versus* real-life behavior of physicians/patients, or between study designs/methods. Moreover, it may provide promising implications for future research.

Practical Implications for Researchers

The third paradigm provides a general research framework that allows the identification of those contextual factors interacting with the drug that have a particular impact on its effectiveness ("drivers of effectiveness").

All contextual factors may not equally impact a potential EEG. For instance, in the study by Schneeweiss et al. [58], several "levels of restriction" were applied to a cohort study population to replicate finding of RCTs, but did not share a similar impact on estimates for effects on statins: the restriction on "incident users" (by excluding prevalent users) had a greater impact than did the restriction on "adherent patients" (by excluding nonadherent patients). Through this example, we may hypothesize that the "incident/prevalent user" characteristic "drives" the effectiveness of statins, whereas "adherence" does not. This hypothesis, which of course would need to be further addressed, has practical implications in the design of an RCT to establish the effectiveness of statins among all potential users, including new users and possible switchers. A practical implication in designing an RCT on statins would be to include both incident and prevalent users, so as to be able to take this factor into account and minimize the risk of an EEG.

Another issue to consider is that drivers of effectiveness might depend on the drug and the disease under investigation. For instance, the choice between a double-blind or open-label design was found to influence antidepressants' effect estimates [41,60] because treatment expectancies on the treatment's effect are high in depression but this might not be the case for other drugs or diseases.

The implications for future research are numerous. The identification of the most impactful contextual factors may help to better design RCTs and if the latter are to be designed in a more "pragmatic" way [10], the emphasis in pragmatism should be put on these impactful contextual factors. Also, the contextual factors that are most likely to cause an EEG need to be identified "case by case," that is, for each disease area.

Study Limitations

First, the present review was not exhaustive but was not aimed to be so. The literature published on the EEG is extensive, and a selection was made using predefined criteria (publication in English, peer-reviewed articles, subjective appraisal of articles' quality). Also, the "snowball method" is thought to identify the most cited articles, usually published in journals in the higher impact factor scale range [62]. A selection bias preventing us from identifying all the existing paradigms on the EEG, however, cannot be excluded.

Second, our study focused on the gap in effectiveness outcomes and did not explore a gap in safety or benefit-risk ratio outcomes. Purposely made, this choice meets the objectives of the GetReal project, allowing us to better frame our research to that end. A gap in safety outcomes, however, may exist exactly in the same way as it does for effectiveness outcomes. The articles reviewed in the present study often discussed the two kinds of outcomes indifferently. For instance, Nallamothu et al. [50] have provided examples for a gap in knowledge in the risk of hyperkalemia in patients taking aldactone, which was underestimated in RCTs. Hernan et al. [63] have investigated the discrepancies of results between cohort studies and RCTs regarding the risk of coronary heart disease in women subject to hormone replacement therapy. The authors have evidenced that this discrepancy could be largely explained by differences in the distribution of time elapsed since menopause (within 10 years of menopause vs. more), with a significant interaction found for this factor. Overall, the results of the present study may certainly apply to safety or benefit-risk outcomes.

Finally, the present review did not explore to which extent the identification of an EEG may have impacted stakeholders decisions. The issue of the EEG was first raised by prescribers and health technology assessment (HTA) bodies, but has now reached the awareness of regulators [64]. The European Medicines Agency and European HTAs now offer parallel scientific advice to manufactures, where HTA needs for real-world data are being voiced early in development [65]. It would be interesting to investigate whether, and to which extent, regulatory decisions have been made in light of the issue of the EEG, that is, using real-life evidence. This is the objective and scope of another GetReal study in which we systematically reviewed phase 3 preauthorization trials that used more "pragmatic" designs (Karcher et al, 2015, submitted manuscript, unpublished data).

Conclusions

Identifying and targeting the contextual factors that have a meaningful impact on effect estimates for medications is a key priority for RCT design permitting better estimates of the effectiveness of drugs in addition to their efficacy. Within the GetReal Consortium [7], several studies are under way to identify the drivers of effectiveness in different therapeutic areas (schizophrenia, Hodgkin's lymphoma, diabetes, etc.). We hope that this review will help future research build on the identified paradigms.

Source of financial support: The work leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking (grant agreement no. 115546), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and in kind contribution from European Federation of Pharmaceutical Industries and Associations (EFPIA) companies.

REFERENCES

[1] Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. BMJ 1999;319:652–3.
[2] Strom B, Kimmel SE. Textbook of Pharmacoepidemiology. John Wiley & Sons, Ltd. Chichester, UK, 2006.
[3] Silverman E. Effectiveness/efficacy difference too often ignored. Manag Care 2013;22:36.
[4] Pagoto SL, Lemon SC. Efficacy vs effectiveness. JAMA Intern Med 2013;173:1262–3.
[5] Eichler HG, Abadie E, Breckenridge A, 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.
[6] The Green Park Collaborative. Review of Evidence Needs for Pharmaceutical Therapies for Alzheimer's Disease. 2012. Available from: www.greenparkcollaborative.org. [Accessed October 10, 2015].
[7] Innovative Medicines Initiative. The GetReal Consortium. 2013. Available from: http://www.imi-getreal.eu/. [Accessed October 10, 2015].
[8] The Food and Drug Administration. Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products. Center for Drug Evaluation and Research, 1998. Available from: http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm078749.pdf. [Accessed October 10, 2015].
[9] The Working Group on Relative Effectiveness. Core principles on relative effectiveness. High level group on innovation and provision of medicines in the EU. Available from: http://ec.europa.eu/DocsRoom/documents/7581?locale=en. [Accessed October 3, 2015].
[10] Center for Medical Technology Policy. Pragmatic phase 3 pharmaceutical trials: recommendations for the design of clinical trials that are more informative for patients, clinicians, and payers. 2010. Available from: http://www.cmtpnet.org/docs/resources/PCT3_EGD.pdf. [Accessed October 30, 2015].
[11] Clinical Trials Transformation Initiative. Large simple trials: facilitating the use of large simple trials. 2012. Available from: http://www.ctti-clinicaltrials.org/what-we-do/investigational-plan/large-simple-trials. [Accessed October 30, 2015].
[12] EUnetHTA. Position Paper on How to Decide on the Appropriate Study Design for Primary Research Arising From HTA reports. 2015.
[13] Mays N, Pope C. Qualitative research in health care: assessing quality in qualitative research. BMJ 2000;320:50–2.
[14] Dixon-Woods M, Sutton A, Shaw R, et al. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. J Health Serv Res Policy 2007;12:42–7.
[15] Banta D. The development of health technology assessment. Health Policy 2003;63:121–32.
[16] Hussey HH. Changing dimensions of medical knowledge: their implications for medical education, research, and practice. J Am Med Assoc 1958;167:40–7.
[17] Lasagna L. The controlled clinical trial: theory and practice. J Chronic Dis 1955;1:353–67.
[18] Cochrane AL. Effectiveness and Efficiency: Random Reflections on Health Services. In: Trust NPH, ed. Royal Society of Medicine Press: London; 1972.
[19] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis 1967;20:637–48.
[20] Boissel JP, Sacks HS, Leizorovicz A, et al. Meta-analysis of clinical trials: summary of an international conference. Eur J Clin Pharmacol 1988;34:535–8.
[21] The Cochrane Collaboration. 2014. Available from: http://www.cochrane.org/. [Accesesd October 30, 2015].
[22] Lowe CU. The consensus development programme: technology assessment at the National Institute of Health. BMJ 1980;280:1583–4.
[23] Howick JH. The Philosophy of Evidence-Based Medicine. Wiley-Blackwell, Chichester, UK, 2011.

[24] Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71–2.
[25] Depp C, Lebowitz BD. Clinical trials: bridging the gap between efficacy and effectiveness. Int Rev Psychiatry 2007;19:531–9.
[26] Lehman AF, Thompson JW, Dixon LB, Scott JE. Schizophrenia: treatment outcomes research–editors' introduction. Schizophr Bull 1995;21:561–6.
[27] Institute of Medicine. Assessing Medical Technologies. N.A. Press, ed. Washington, DC: National Academy of Sciences, 1985.
[28] Timbie JW, Fox DS, Van Busum K, Schneider EC. Five reasons that many comparative effectiveness studies fail to change patient care and clinical practice. Health Aff (Millwood) 2012;31:2168–75.
[29] Sullivan M. The new subjective medicine: taking the patient's point of view on health care and health. Soc Sci Med 2003;56:1595–604.
[30] Sabate E. Adherence to Long-Term Therapies: Policy for Action. Meeting Report, WHO Adherence Project, 4-5 June 2001.
[31] Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011;86:304–14.
[32] Scott J. Using Health Belief Models to understand the efficacy-effectiveness gap for mood stabilizer treatments. Neuropsychobiology 2002;46(Suppl. 1):13–5.
[33] Morrison VL, Holmes EA, Parveen S, et al. Predictors of self-reported adherence to antihypertensive medicines: a multinational, cross-sectional survey. Value Health 2015;18:206–16.
[34] Bennell KL, Egerton T, Bills C, et al. Addition of telephone coaching to a physiotherapist-delivered physical activity program in people with knee osteoarthritis: a randomised controlled trial protocol. BMC Musculoskelet Disord 2012;13:35.
[35] Demonceau J, Ruppar T, Kristanto P, et al. Identification and assessment of adherence-enhancing interventions in studies assessing medication adherence through electronically compiled drug dosing histories: a systematic literature review and meta-analysis. Drugs 2013;73:545–62.
[36] Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med 2005;353:1209–23.
[37] El-Serag HB, Talwalkar J, Kim WR. Efficacy, effectiveness, and comparative effectiveness in liver disease. Hepatology 2010;52:403–7.
[38] Thornicroft G, Ruggeri M, Goldberg D. Improving Mental Health Care: The Global Challenge. Hoboken, NJ: Wiley-Blackwell, Chichester, UK, 2013.
[39] Unutzer J, Katon W, Sullivan M, Miranda J. Treating depressed older adults in primary care: narrowing the gap between efficacy and effectiveness. Milbank Q 1999;77(225–56):174.
[40] Bauer MS, McBride L, Shea N, et al. Impact of an easy-access VA clinic-based program for patients with bipolar disorder. Psychiatr Serv 1997;48:491–6.
[41] Naudet F, Maria AS, Falissard B. Antidepressant response in major depressive disorder: a meta-regression comparison of randomized controlled trials and observational studies. PLoS One 2011;6:e20811.
[42] Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol 2003;3:28.
[43] Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet 2002;359:248–52.
[44] McMahon AD. Study control, violators, inclusion criteria and defining explanatory and pragmatic trials. Stat Med 2002;21:1365–76.
[45] Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. J Clin Epidemiol 1996;49:749–54.
[46] Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. Lancet 2008;372:2152–61.
[47] Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med 2008;5:e67.
[48] Hordijk-Trion M, Lenzen M, Wijns W, et al. Patients enrolled in coronary intervention trials are not representative of patients in clinical practice: results from the Euro Heart Survey on Coronary Revascularization. Eur Heart J 2006;27:671–8.
[49] Lewis JH, Kilgore ML, Goldman DP, et al. Participation of patients 65 years of age or older in cancer clinical trials. J Clin Oncol 2003;21:1383–9.
[50] Nallamothu BK, Hayward RA, Bates ER. Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. Circulation 2008;118:1294–303.
[51] Simon G, Wagner E, Vonkorff M. Cost-effectiveness comparisons using "real world" randomized trials: the case of new antidepressant drugs. J Clin Epidemiol 1995;48:363–73.
[52] Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine". Am J Med 1997;103:529–35.
[53] Steckler A, McLeroy KR. The importance of external validity. Am J Public Health 2008;98:9–10.
[54] 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;290:1624–32.
[55] Chakravarty EF, Fries JF. Science as experiment; science as observation. Nat Clin Pract Rheumatol 2006;2:286–7.
[56] Atar D, Ong S, Lansberg PJ. Expanding the evidence base: comparing randomized controlled trials and observational studies of statins. Am J Ther 2015;22:e141–50.
[57] Grootendorst DC, Jager KJ, Zoccali C, Dekker FW. Observational studies are complementary to randomized controlled trials. Nephron Clin Pract 2010;114:c173–7.
[58] Schneeweiss S, Patrick AR, Sturmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. Med Care 2007;45(10, Suppl. 2):S131–42.
[59] Ankarfeldt MZ, Angquist L, Stocks T, et al. Body composition, dietary protein and body weight regulation: reconciling conflicting results from intervention and observational studies? PLoS One 2014;9:e101134.
[60] Chassang S, Snowberg E, Seymour B, Bowles C. Accounting for behavior in treatment effects: new applications for blind trials. PLoS One 2015;10:e0127227.
[61] Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol 2009;62:464–75.
[62] Falagas ME, Zarkali A, Karageorgopoulos DE, et al. The impact of article length on the number of future citations: a bibliometric analysis of general medicine journals. PLoS One 2013;8:e49476.
[63] Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology 2008;19:766–79.
[64] Eichler HG, Bloechl-Daum B, Abadie E, et al. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. Nat Rev Drug Discov 2010;9:277–91.
[65] European Medicines Agency. Best practice guidance for Pilot EMA HTA Parallel Scientific Advice procedures. 2014. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2014/05/WC500166226.pdf. [Accessed October 30, 2015].