Comparative Effectiveness Research/Health Technology Assessment (HTA)

The “Efficacy-Effectiveness Gap”: Historical Background and Current Conceptualization

Clementine Nordon, MD PhD1,*, Helene Karcher, PhD2, Rolf H.H. Groenwold, PhD3, Mikkel Zöllner Ankarfeldt, PhD4, Franz Pichler, PhD5, Helene Chevrou-Severac, PhD6, Michel Rossignol, MD, PhD7, Adeline Abbe, MSc8, Lucien Abenhaim, MD, PhD2, Australia; 6Takeda Pharmaceuticals International, Glattpark-Opfikon, Switzerland; 7Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada; 8Sanofi R&D, Chilly Mazarin, France

on behalf of the GetReal consortium

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.

Copyright © 2016, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

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...
Table 1 - Search algorithm on the “efficacy-effectiveness gap”.

<table>
<thead>
<tr>
<th>Efficacy</th>
<th>Effectiveness</th>
<th>Real life</th>
<th>CER</th>
<th>Concept</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 CER</td>
<td>CER [Title/Abstract]</td>
<td>CER [Title/Abstract]</td>
<td>CER [Title/Abstract]</td>
<td>CER [Title/Abstract]</td>
<td>CER [Title/Abstract]</td>
</tr>
</tbody>
</table>

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 synthesize 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 synthesizing 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 synthesize 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
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 biological 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

---

**Table 2 - Summary of the three paradigms identified.**

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Description</th>
<th>Themes encompassed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The EEG is related to real-life characteristics of the health care system</td>
<td>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</td>
<td>• In routine practice, the physicians’ “behavior” regarding medical guidelines and dissemination of knowledge is not optimal&lt;br&gt;  • In routine practice, the patients’ adherence is not optimal&lt;br&gt;  • In routine practice, there are access barriers to health care resources&lt;br&gt;  • 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&lt;br&gt;  • Concept of pragmatism: RCTs’ lack of generalizability; any direct dissemination of evidence coming from clinical trial into clinical practice is inadequate&lt;br&gt;  • Some contextual factors are (significantly) interacting with the drug’s biological effect (“drivers of effectiveness”)&lt;br&gt;  • An imbalance in the distribution of these factors between efficacy and effectiveness studies may cause an EEG</td>
</tr>
<tr>
<td>2. The EEG is related to an issue of the method used to measure the drug’s effect</td>
<td>Efficacy and effectiveness studies use different study designs and design parameters, hence the EEG</td>
<td></td>
</tr>
<tr>
<td>3. The EEG is related to an issue of complex interaction</td>
<td>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</td>
<td></td>
</tr>
</tbody>
</table>

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