Methodological Considerations in the Assessment of Effectiveness of Homeopathic Care: A Critical Review of the EPI3 Study

Yola Moride\textsuperscript{1,2,3}\textsuperscript{\dagger}

\textsuperscript{1} Faculty of Pharmacy, Université de Montréal, Montreal, Quebec, Canada
\textsuperscript{2} Center for Pharmacoepidemiology and Treatment Science, Rutgers, The State University of New Jersey, New Brunswick, New Jersey, United States
\textsuperscript{3} YolaRX Consultants, Montreal, Quebec, Canada

Homeopathy 2022;111:147–151.

\textbf{Abstract}

\textbf{Background} EPI3 is an observational study of a representative sample of general practitioners (GPs) and patients in France, demonstrating that patient characteristics differ according to the prescribing preferences of their GPs for homeopathy. For selected conditions (musculoskeletal disorders, sleep disorders, anxiety/depression, upper respiratory tract infections), progression of symptoms and adverse events over follow-up in the homeopathy preference group did not significantly differ from other practice preferences, but there was a two-fold to four-fold lower usage of conventional medicines. The EPI3 study’s validity was challenged due to absence of head-to-head comparison of medicines to conclude on a causal association between homeopathy and outcomes.

\textbf{Methods} A critical review of the nine EPI3 publications was conducted, focusing on generalizability, selection bias, outcome measurements and confounding.

\textbf{Results} The conceptual framework of EPI3 rests on a systemic construct, i.e., the homeopathic treatment concept assessed using the type of GP prescribing preference, taking into account the clinical, human and social aspects. The enrollment process enhanced the generalizability of findings. Validated instruments for outcome measurements were used for three conditions, and control of confounding was rigorous.

\textbf{Conclusion} EPI3 was conducted according to best practices. Homeopathy prescribing preference met specific patient needs with less use of conventional medicines and without an apparent loss in therapeutic opportunity.
GP practices on a random practice day (maximum 15 patients per site). Study patients who consulted their GP for either musculoskeletal disorders (n = 1,153), sleep disorder (n = 346), anxiety/depression (n = 710) or upper respiratory tract infections (URTIs) (n = 518) were enrolled in a prospective cohort study, with a follow-up of 12 months (with contacts at baseline, 1 month, 3 months and 12 months), that aimed at comparing the three prescribing preference groups on the basis of clinical and symptomatic outcomes, patient-reported outcomes (quality of life), drug utilization and adverse events that indicated loss of therapeutic opportunity. The study found slight differences in the sociodemographic characteristics of patients who are followed in the homeopathy preference group, consisting of a greater proportion of females and individuals with higher education, but no differences in co-morbidities and quality of life. However, they had markedly healthier lifestyle and positive attitude toward complementary and alternative medicines.

In each of the four disease cohorts, patients followed in the homeopathy preference practices used fewer conventional medicines than those followed in the conventional practices, ranging from a more than two-fold decrease in the use of antibiotics in the URTI cohort and in the use of NSAIDs in the musculoskeletal disease cohort to a four-fold decrease in psychotropic drug use in the sleep disorder and anxiety/depression cohorts. Differences between practice preferences regarding symptoms improvement or progression were not statistically significant. Criticisms were raised by the French Health Authority (Haute Autorité de Santé [HAS]) regarding the validity of the EPI3 study: namely the absence of a head-to-head comparison of medicines to conclude on a causal association between homeopathy and outcomes, as well as the non-comparability of patients followed in the different types of practice preferences.

The present Commentary briefly addresses the following: what evidence has been generated by the EPI3 study with regard to the study findings, the quality of evidence including methodological robustness, and the generalizability of findings?

Methods

A critical review of the methodological quality of the EPI3 study was conducted. Sources of information were the nine papers published in the literature by the co-investigators. Three follow-up studies consisting of economic evaluations were not considered in the evaluation given that they involved secondary analyses of existing data, were not planned a priori, and were not conducted by the original study investigators. The review was conducted against the methodological criteria described in the Guidelines for Good Pharmacoepidemiology Practices (version 4, 2015), the ENCePP Guide on Methodological Standards in Pharmacoepidemiology (Version 8, 2020), and the Joanna Briggs Institute (JBI) Critical Appraisal Tools. More specifically, the following four dimensions of study quality were assessed: (1) generalizability of study findings to the French population; (2) validity of the selection process for the enrollment of GPs and patients in the study; (3) quality of data collected, including validity of instruments to measure changes in clinical and humanistic outcomes over time; (4) analytical methods applied to take into account the non-comparability of patients followed in the various prescribing preference practices (homeopathy, conventional, mixed).

Results

Conceptual Framework

Health care practice needs to be based on robust theory that guides efforts in practice and research. Evaluation is key to inform best practices, using proximal and/or distal outcomes. Although not explicitly stated in any of the EPI3 papers, the study reports on an evaluation conducted after the GPs selected and implemented their prescribing preference. The main construct of the conceptual framework that was evaluated was the effectiveness of the homeopathic treatment concept assessed using the type of GP prescribing preference (homeopathy, strictly conventional, or mixed), taking into account the clinical, human and social aspects. Proximal outcome consisted of conventional drug usage during follow-up, whilst distal outcomes included symptoms progression, humanistic outcomes (quality of life) and loss of therapeutic opportunity. It is important to mention that the framework does not include the evaluation of the efficacy of the homeopathic medicinal product compared to conventional medicine as one of the constructs; hence, the absence of head-to-head comparison between products. A more explicit description of the conceptual framework of the EPI3 study in the original papers would have been helpful.

Generalizability of Study Findings

The quality of evidence generated by the cross-sectional study on patient characteristics according to type of prescribing preference relies on the appropriateness of the sampling strategy. In EPI3, the sampling unit was the GP and the unit of observation was the patient. A two-stage sampling process was used whereby, first, a random sample of GPs stratified by prescribing preference (homeopathy-certified vs. non-homeopathy-certified) was drawn from the French national directory of physicians. The sample was stratified according to practice preference (homeopathy, strictly conventional, mixed). Second, a random practice day was selected for each sampled GP in order to enroll patients who attended the clinic on that day (up to a maximum of 15 per GP). Unlike many population samples based on primary data collection, the EPI3 study benefited from the availability of the French national directory of physicians, which provided a well-defined roster of practicing GPs in the country, with selected characteristics such as region and practice type (homeopathy vs. no homeopathy). Furthermore, participating patients were compared to non-participating eligible patients on the basis of gender, age, length of time attending the GP practice, type of health insurance and main reason for consultation, which allowed for the calibration of the patient sample. Results on the
characteristics of patients attending the three types of practices based on their prescribing preferences for homeopathy (homeopathy, strictly conventional, mixed) were thus weighted according to patient and GP distribution, using CALMAR (a well-known procedure in demographic studies designed to enhance the representativeness of results to the source population). Generalizability of findings was thus an important strength of the EPI3 study because it aimed at describing practice and patient characteristics.

**Selection Process**
Selection bias is a threat to observational studies conducted using primary data collection as it is necessary to sample and contact participants. As mentioned in the previous section, great care was taken in the EPI3 study to obtain a representative sample of GPs and patients. In addition, at the time of patient enrollment, broad objectives were presented to potential participants without disclosing the study hypothesis; this methodological feature was important in order to avoid influencing participation and the self-reported information. Furthermore, unique to the EPI3 study was the creation of a registry of patients who declined participation based on data provided by the GP (patient age, gender, socio-demographics, education, smoking, body mass index, employment status, lifestyle, history of hospitalization in the previous 12 months, whether the GP was the regular physician or not, number of GP consultations in the past year, quality of life [SF-12], attitude and beliefs regarding complementary and alternative medicines [CAMBI score]); and 2) self-reported data at each follow-up (history of drug utilization using the standardized Progressive Assisted Backward Active Recall [PABAR]). One of the advantages of patient interviews is the availability of information on over-the-counter drug usage. This instrument has been validated against prescription data only for cardiovascular drugs and drugs used for musculoskeletal disorders, the latter being one of the cohorts of interest in EPI3. Because of over-the-counter and use on an as-needed basis, agreement between self-report and physician’s prescription was shown to be lower than for cardiovascular drugs. In addition, for each cohort, disease-specific instruments were used to enroll patients at baseline and monitor disease progression and symptoms over time as study outcomes. The robustness of the outcome measures varied between the cohorts, and the most robust were those used in the musculoskeletal disease cohort, whereby functional status was measured using the Roland-Morris score (back pain), Quick Dash (upper limbs), and Lequesne (lower limbs). Scores were standardized to 100 and improvement was defined by an increase of 12.5 points relative to baseline, this threshold having been validated in the literature.

In the anxiety/depression cohort, the Hospital Anxiety and Depression Scale (HADS) was used to ascertain cases, using a threshold of 9. This threshold has been validated in primary care against DSM-III-defined psychiatric morbidity for use as a case finder. However, study outcome of improvement in symptoms, defined by a HADS score less than 9, does not appear to have been validated to measure improvement, though both are related. In EPI3, the threshold was thus determined a priori based on clinical expert assessment as opposed to a formal validation of the HADS scale to detect improvement. Because the mean HADS score at baseline was similar across the three prescribing preference groups (ranging from 11.8 to 12.0), one would not expect a differential misclassification of outcome across GP types. In the sleep disorder cohort, the Pittsburgh sleep quality index (PSQI) ≥1 was used to find cases as well as to define outcome (persistence of sleep disorder over the follow-up). The PSQI has been validated in the literature. In addition, self-reported occurrence of any injury resulting from a fall, motor vehicle collision, sports or employment over the 12-month follow-up was another study outcome. In the URTI cohort, outcome consisted of self-reported change in URTI symptoms (cleared, much improved, slightly improved, no change, or worsened) at one month and associated infections (otitis or sinusitis) in the 12-month follow-up.

**Data Collection Tools**
The EPI3 study was based on primary data collection, using questionnaires to GPs and patients. Across the four cohort studies, two types of data were collected from patients: 1) self-reported information collected at enrollment (i.e., socio-demographics, education, smoking, body mass index, employment status, lifestyle, history of hospitalization in the previous 12 months, whether the GP was the regular physician or not, number of GP consultations in the past year, quality of life [SF-12], attitude and beliefs regarding complementary and alternative medicines [CAMBI score]); and 2) self-reported data at each follow-up (history of drug utilization using the standardized Progressive Assisted Backward Active Recall [PABAR]). One of the advantages of patient interviews is the availability of information on over-the-counter drug usage. This instrument has been validated against prescription data only for cardiovascular drugs and drugs used for musculoskeletal disorders, the latter being one of the cohorts of interest in EPI3. Because of over-the-counter and use on an as-needed basis, agreement between self-report and physician’s prescription was shown to be lower than for cardiovascular drugs. In addition, for each cohort, disease-specific instruments were used to enroll patients at baseline and monitor disease progression and symptoms over time as study outcomes. The robustness of the outcome measures varied between the cohorts, and the most robust were those used in the musculoskeletal disease cohort, whereby functional status was measured using the Roland-Morris score (back pain), Quick Dash (upper limbs), and Lequesne (lower limbs). Scores were standardized to 100 and improvement was defined by an increase of 12.5 points relative to baseline, this threshold having been validated in the literature.

In the anxiety/depression cohort, the Hospital Anxiety and Depression Scale (HADS) was used to ascertain cases, using a threshold of 9. This threshold has been validated in primary care against DSM-III-defined psychiatric morbidity for use as a case finder. However, study outcome of improvement in symptoms, defined by a HADS score less than 9, does not appear to have been validated to measure improvement, though both are related. In EPI3, the threshold was thus determined a priori based on clinical expert assessment as opposed to a formal validation of the HADS scale to detect improvement. Because the mean HADS score at baseline was similar across the three prescribing preference groups (ranging from 11.8 to 12.0), one would not expect a differential misclassification of outcome across GP types. In the sleep disorder cohort, the Pittsburgh sleep quality index (PSQI) ≥1 was used to find cases as well as to define outcome (persistence of sleep disorder over the follow-up). The PSQI has been validated in the literature. In addition, self-reported occurrence of any injury resulting from a fall, motor vehicle collision, sports or employment over the 12-month follow-up was another study outcome. In the URTI cohort, outcome consisted of self-reported change in URTI symptoms (cleared, much improved, slightly improved, no change, or worsened) at one month and associated infections (otitis or sinusitis) in the 12-month follow-up.

**Statistical Analysis**
In order to compare study outcomes (conventional drug use, symptoms progression) across the prescribing preference types, it was necessary to adjust for differences in patient characteristics. Those characteristics included gender, age group, body mass index, smoking habit, alcohol consumption, physical activity, education, employment, complementary health insurance, CAMBI (Complementary and Alternative Medicine Beliefs Inventory) score, as well as baseline value of disease-specific assessment scores. Propensity scores were derived with those covariates, using multivariate logistic regression. Two sets of propensity scores were derived, indicating the probability of patients to belong to homeopathy versus conventional practice type, and mixed versus conventional practice type, conditional on all other...
baseline variables. Propensity scores have become the cornerstone of comparative safety and effectiveness research and, when used for matching, are referred to as pseudo-randomization. For binary outcome variables (e.g., disease progression, conventional drug use), multivariate logistic regression analyses were conducted, adjusting for baseline disease characteristics, age, gender and propensity scores. For continuous outcome variables (e.g., functional scores), multivariate ANOVA for repeated measures was used. Non-independence of patients within a given GP type, as well as autocorrelation between responses to the four consecutive interviews, were addressed through generalized estimating equations in the multivariate models. Furthermore, stratification by incident versus prevalent disease status at baseline was conducted.

**Conclusion**

EPI3 is a large-scale epidemiologic study that confirmed heterogeneity in the characteristics of patients who consult GPs that differ according to their prescribing preferences for homeopathy and complementary medicine. The EPI3 study was based on a systemic construct defined by the type of GP prescribing preference, and was not designed to perform head-to-head comparisons of individual medicines. Owing to careful methods of sampling, findings are highly generalizable to the French population. Overall, evidence generated by the EPI3 study is robust, especially for musculoskeletal diseases, since validated instruments were used for both case ascertainment and study outcomes. As the EPI3 study was non-randomized, the non-comparability of patients followed in the various types of GP practices was addressed through advanced analytic techniques consisting of propensity scores, which are considered to be pseudo-randomization and part of good pharmacoepidemiology practice.

The EPI3 study showed that patients followed in homeopathic preference practice use fewer conventional medicines and without statistically significant differences in symptoms progression or adverse events compared to the other prescribing preferences, thereby indicating no loss of therapeutic opportunity. The EPI3 study highlighted differences between practice types regarding patient characteristics and behaviors. Thus, given those specificities in patient populations, homeopathic practice met specific patient needs.

- Outcomes were measured using validated scales and instruments for three conditions of interest (musculoskeletal disorders, sleep disorder, anxiety/depression), though only the one for musculoskeletal disease had been validated as a measure of symptoms improvement.
- As the EPI3 study was non-randomized, the non-comparability of patients followed in the various types of GP practices was addressed through advanced analytic techniques consisting of propensity scores.

**Funding**

This review was sponsored by Boiron.

**Conflict of Interest**

Yola Moride is a consultant for the pharmaceutical industry, and played no part in the original EPI3 study or its associated discussions. Critical review was conducted at arm’s length, whereby the sponsor did not influence the assessments.

**Highlights**

- The conceptual framework for the EPI3 study rests on a systemic construct (GP prescription preference) as opposed to head-to-head comparison of individual medicines.
- Through the enrollment process that was used, the study population was representative of the French population, and selection bias was minimized through adequate study procedures and a registry of non-participants.
- The EPI3 study was based on primary data collection, provided by GPs and self-reported by participants.

**References**

20 Coste J, Le Parc JM, Berge E, Delecroeuillerie G, Paolaggi JB. French validation of a disability rating scale for the evaluation of low back pain (EIFEL questionnaire). Rev Rhum 1993;60:335–341