Published in final edited form as:

Clin Pharmacol Ther. 2019 May ; 105(5): 1156–1163. doi:10.1002/cpt.1210.

Using Real-World Data to Extrapolate Evidence From Randomized Controlled Trials

Shirley V. Wang¹, Sebastian Schneeweiss¹, Joshua J. Gagne¹, Thomas Evers², Christoph Gerlinger^{3,4}, Rishi Desai¹, and Mehdi Najafzadeh¹

¹Department of Medicine, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

²Market Access, Bayer AG, Wuppertal, Germany

³Statistics and Data Insights, Bayer AG, Berlin, Germany

⁴Obstetrics and Gynecology, Saarland University, Homburg/Saar, Germany.

Abstract

Randomized controlled trials (RCTs) provide evidence for regulatory agencies, shape clinical practice, influence formulary decisions, and have important implications for patients. However, many patient groups that are major consumers of drugs are under-represented in randomized trials. We review three methods to extrapolate evidence from trial participants to different target populations following market approval and discuss how these could be implemented in practice to support regulatory and health technology assessment decisions. Although these methods are not a substitute for less restrictive pre-approval RCTs or rigorous observational studies when sufficient data are available in the post-approval setting, they can help to fill the evidence gap that exists in the early marketing period. Early evidence using real-world data and methods for extrapolating evidence should be reported with clear explanation of assumptions and limitations especially when used to support regulatory and health technology assessment decisions.

Randomized controlled trials (RCTs) provide evidence for regulatory agencies, shape clinical practice, influence formulary decisions, and have important implications for patients. In the United States, when a new drug enters the market, evidence of its benefits and risks is often based solely on data from RCTs conducted for regulatory agency approval. For decades, it has been recognized that many patient groups, such as older adults, racial minorities, pregnant women, and others, have been under-represented in RCTs although also being major consumers of drugs.^{1,2} For example, although ischemic heart disease is prevalent in older adults, over 50% of trials registered at clinicaltrials.gov between January

Correspondence: Shirley V. Wang (Swang1@bwh.harvard.edu).

CONFLICT OF INTEREST

S.V.W. receives salary support from investigator initiated grants to Brigham and Women's Hospital from Bayer and Novartis and is a consultant to Aetion, Inc., a software company for unrelated work. J.J.G. has received salary support from investigator initiated grants to Brigham and Women's Hospital from Eli Lilly and Company and Novartis Pharmaceuticals Corporation and is a consultant to Aetion, Inc. and Optum, Inc., all for unrelated work. S.S. is a consultant to WHISCON, LLC, and to Aetion, Inc., a software manufacturer of which he also owns equity. He is principal investigator of investigator-initiated grants to the Brigham and Women's Hospital from Bayer, Genentech, and Boehringer Ingelheim. T.E. and C.G. are employed by Bayer AG.

2006 and January 2016 explicitly excluded the elderly.³ Similarly, even when not explicitly excluded, women remain under-represented, in part due to fertility-related exclusions. Although diabetes is a chronic condition with significant morbidity that affects men and women differently,⁴ nearly 60% of trials for type 2 diabetes medications registered at clinicaltrials.gov had at least one fertility-related exclusion criterion; 7% had exclusions for women with child-bearing potential.⁵ Patients with complex comorbidities are also often excluded from trials. One report found that although chronic kidney disease is associated with a high risk of death from coronary artery disease and that effectiveness of treatment can be different in this population, 75% of registered trials for drugs to treat coronary artery disease excluded patients with chronic kidney disease.⁶ Such exclusions limit the generalizability of evidence from trials. In spite of the discordance between the patients in whom trials are conducted and the patients in whom these therapies are used in practice, the focus of RCTs on highly selected populations is unlikely to change in the near future. It is important to understand how to make better use of available data to inform regulators, payers, and clinical practice in a timely manner.

Although RCT evidence has strong internal validity, it may not be reflective of treatment benefit and risk in all patient populations that will actually be treated in practice. The effect of drugs can be different under real-world conditions, where there will often be lower adherence to therapy than observed in trial participants. There may also be differences in treatment effect across patient populations that were and were not included in the trials due to differences in patient characteristics. Furthermore, in populations in which the relative efficacy of treatments is uniform, the additive effect will be heterogeneous and vice versa. This means that in trials with a homogeneous relative effect, the additive effect (risk difference) can vary substantially for patients with different baseline risks.⁷

Large, administrative healthcare databases, which prospectively record information on healthcare utilization, reflect real-world treatment patterns among diverse populations.^{8–10} RCTs and healthcare databases are often used to address different aspects of drug effects, with the former typically focused on determining efficacy and the latter on evaluating “real-world” effectiveness and safety.^{11,12} However, evidence of efficacy from RCTs could be generalized (“standardized”) to a target population with a different distribution of baseline characteristics or that includes patients who would have been excluded from trials but that use the treatments in the real world. Making fuller use of individual-level RCTs and electronic healthcare data from patients treated in routine care could enhance pre-approval or postapproval evidence in a meaningful way. Prior studies have highlighted the variation in effect of treatments for patients within randomized trials, showing that benefit on the additive scale (risk difference) is often concentrated among older, sicker participants who are at highest risk of the outcome.¹³ The skewed distribution means that more than half of patients can have no net benefit, yet the average effect in the population is beneficial. This further implies that when the distribution of baseline risk for participants in the trial and patients treated in routine care differ, the average benefit on the risk difference scale in the excluded routine care population will also differ. Methods to generalize the baseline characteristics of the RCT population to match that of patients treated in routine care can facilitate generation of early evidence for the effectiveness and safety of treatments in excluded populations.

In this article, we review three methods to generalize or extrapolate evidence from participants in a phase III trial to a target population following market approval and discuss how these could be implemented in practice to support regulatory and health technology assessment decisions. The strategies we discuss include (i) re-weighting (standardization) methods,¹⁴ (ii) cross-design synthesis,¹⁵ and (iii) discrete event simulation.¹⁶

APPROACH

As a motivating example, we focus on a fictional phase III trial investigating the efficacy of a new drug, fantastistatin, compared to the existing standard of care, normastatin, on risk of major adverse cardiac events (MACEs). Eligibility criteria for the pivotal phase III trial for fantastistatin included restrictions to participants who were < 65 years of age and who were hospitalized with a myocardial infarction (MI) within 6 months prior to enrollment. However, the patients who might use the treatment in practice (i.e., the target population in which one might be interested in extending trial results) includes those over 65 years of age as well as patients who may have had an MI > 6 months before treatment initiation. The trial also enrolled predominantly male patients who were generally healthier than the target population. A lower percentage of trial participants had diabetes and hypertension than typically observed in patients with prior MI.

The distribution of baseline characteristics of participants in this fictitious phase III trial differs from the distribution in the target population likely to receive treatment following market entry. As shown in Figure 1a, the majority of trial participants are between 30 and 50 years of age and enroll in the trial 2–4 months after hospitalization for MI. In contrast, the target population that receives treatment in routine care is predominantly 50–85 years old and most patients do not initiate treatment until 6–12 months after an MI.

If treatment effect varies based on patient characteristics—for example, a stronger additive or relative effect in sicker patients—then the average effect observed in trial participants may not reflect the average effect in the target population. There is interest in understanding how the effect of fantastistatin may differ for patients who are older at the time of initiation of therapy as well as patients who initiate therapy > 6 months after an MI. Regulatory and health technology assessment agencies would like to use healthcare databases to generate early evidence that can be evaluated in parallel with the pivotal RCT findings to inform early regulatory or health technology assessment decisions.

REVIEW OF METHODS TO GENERALIZE OR EXTRAPOLATE EVIDENCE FROM A PHASE III TRIAL TO A REAL-WORLD POPULATION

Reweighting (standardization) methods

Reweighting or “standardization” techniques are well-established epidemiologic methods that could be applied to individual-level RCT and observational data as a means to enhance understanding of drug effectiveness or safety.^{7,14,17,18} These methods estimate measures of occurrence (such as rates) or association (such as hazard ratios) from the RCT after changing the distribution of baseline characteristics to match those in a target population.

Standardized estimates can be obtained without access to individual level data from the trial or real-world patients if the number of relevant baseline characteristics is small. For example, if age were the only characteristic that modified the hazard ratio for treatment efficacy in the trial, the trial hazard ratios in each age group is known, and the distribution of age in the target population is known, then there would be no need for individual level data to estimate a hazard ratio age-standardized to the target population. When there are multiple baseline characteristics of interest and individual level data are not available, standardization on all of these characteristics simultaneously to the target population would require estimates for the target measure of occurrence or association within all potential subgroup combinations as well as the joint distributions of the subgroups in the target population. This information is often not available, making it difficult to standardize estimates across numerous baseline characteristics without access to individual level data. With access to individual level data in both the trial and an observational healthcare data source, the trial population can be reweighted such that the distribution of baseline characteristics resembles that of a target population following market entry. In order to calculate these weights, one can take these steps:

1. Identify the target population in an observational data source. To the extent possible, use the same inclusion/exclusion criteria and measure the same baseline characteristics as measured in the trial. Focus on identifying patients in observational data who would qualify for the trial.
2. Estimate a propensity score (PS) via a logistic regression predicting whether an individual is a trial participant or identified in observational data. The PS is the predicted probability of being a trial participant conditional on the covariates. The dependent variable is whether the patient was a trial participant or was identified in observational data. The predictors should include baseline covariates that are measured in the trial and the observational data. These scores can be estimated by fitting models after appending individual level trial and observational data together or through distributed regression¹⁹ methods if data access agreements do not permit the two datasets to be stored on the same computing platform.
3. Calculate standardized morbidity ratio weights by assigning trial participants a weight equal to the PS odds (e.g., $PS / (1 - PS)$).²⁰
4. Estimate the average treatment effect in the reweighted trial population. This is the estimated average effect in the target population.

Creation of a dataset including the target population and baseline characteristics from an observational data source can take place while the RCT is enrolling. Trial results can be reweighted to match the target population as soon as the dataset for the primary trial analysis is ready. In order to create PS weights, individual level trial data and observational data must be available simultaneously in the same location or methods for distributed regression must be implemented.²¹

In Figure 1b, the fictional RCT population in our case study is reweighted so that the distribution of trial participant characteristics within the dashed box resembles the

distribution observed in a routine care population that would have met trial eligibility criteria. The reweighted trial data are more heavily tilted toward older patients (mid-50s vs. 40s) and have greater lags between MI and treatment initiation than the unweighted data from trial participants (4–6 months vs. 2–4 months). The average efficacy in this circumscribed target population can then be estimated using the RCT data reweighted so that the distribution of baseline characteristics in trial participants resemble the distribution observed in routine care. This approach has been used in various applications, including reweighting data from human immunodeficiency virus trial participants to estimate highly active antiretroviral therapy efficacy in the general US human immunodeficiency virus infected population¹⁷ and reweighting data from a school-based intervention trial to estimate the effect of the intervention across schools in Maryland.¹⁸

Extrapolation using cross-design synthesis

Sometimes, a trial population does not include members from a potential target population because of restrictive exclusion criteria. Although extrapolation should be conducted and interpreted with caution, and clear understanding of the assumptions being made, such approaches can sometimes help to provide an early assessment of how effects could vary in populations not included in trials.

Cross-design synthesis is a method that borrows information from observational data and combines it with trial data to extrapolate trial estimates to populations that were excluded from participation.^{15,22} While the pivotal trial is enrolling participants, a cohort of users of the comparator treatment in the trial can be identified in observational healthcare data. When the trial is complete and ready for analysis, this cohort can be used to facilitate extrapolation of RCT findings.

There are numerous potential algorithms that can be used to combine information from real-world observational data and trials. For example, the algorithm could extrapolate based on the relative change in incidence rate across different levels of exclusion criteria, or the algorithm could be a multivariable model predicting rates of the outcome for patients excluded from trials.

In our fictional case study, exclusion criteria restricted the trial population to younger age groups and no more than a 6-month lag between prior MI and initiation. The goal was to extrapolate the results from the trial to generate estimates of efficacy for fantastistatin compared to normostatin for a target population in routine care that includes older patients and patients with up to a 12-month lag between MI and treatment initiation (Figure 1c).

We walk through one example of a simple algorithm based on the relative change in rate of MACEs across strata of an exclusion criterion observed in initiators of normostatin in observational data to extrapolate results. Prior studies have implemented more complex cross-design synthesis approaches to extrapolate evidence for diverse applications, including effectiveness of different routes for insulin on blood glucose regulation,¹⁵ efficacy of an angiotensin converting enzyme inhibitor vs. placebo on prevention of heart failure hospitalization or death,²³ and others.^{24,25}

In our fictional example, initiators of the trial comparator, normostatin, can be identified in observational healthcare data, however, fantastistatin is not yet on the market. Rates of MACE following fantastistatin and normostatin initiation for individuals who were excluded from participation in the trial can be extrapolated based on trends in rates for normostatin over values of the trial exclusion criterion that are observed in routine care patients (e.g., trends in rates with increasing age). The extrapolated estimates for rate of events in normostatin and fantastistatin users can then be used to derive estimates of absolute or relative effect in populations that were excluded from the trials.

Steps to implement cross-design synthesis for this example include:

1. Identify the target population (e.g., patients under AND over 65 years of age)
2. Choose an algorithm that uses observed trends in rate of MACEs with increasing age for patients who initiate normostatin in the observational data to extrapolate rates to older patients who were excluded.
3. When feasible, check to see if trends in rate of MACEs with increasing age for normostatin initiators < 65 years in observational data mirror trends in trial participants < 65 years who initiated normostatin. If there are gross deviations, then cross-design synthesis should not be used.
4. Use the algorithm to extrapolate rates of MACEs in both arms of the RCT to estimate the expected rates for each group in older patients who were excluded from the RCT.
5. Weight the observed results for the RCT and extrapolated results for strata that were excluded from the RCT according to the target population distribution (see method A) for estimated average efficacy in the target population. Results can also be provided for relevant age strata within the target population.

In our fictional case study, the rate of MACEs is fairly constant for normostatin initiators between 40 and 60 years of age and increases exponentially at older ages, however, this nonlinear increase is not observed in the trial data because older patients are excluded (Figure 2). The rate of MACEs per 1,000 person-years for normostatin initiators in routine care is 1.02 times higher for patients aged 60–64 years than patients aged 55–59 years at the time of initiation (Table 1). The relative change in rate of MACEs increases for older age groups; the rate is 1.05 times higher for patients aged 65–69 years than patients 60–64 years, and 1.10 times higher for patients 70–74 years than patients aged 65–69 years. The observed change in rate of MACEs for normostatin initiators for age groups under 65 is similar for patients in routine care and trial participants. Because fantastistatin has not yet entered the market, there is no way to evaluate whether the trend in MACEs observed in RCT participants would be mirrored in routine care patients. The exponential change in rate of MACEs for normostatin initiators older than 65 years in routine care patients can be applied to extrapolate rates of MACEs for normostatin and fantastistatin in older adults had they been allowed to participate in the trial. Similarly, the observed change in rate of MACE for normostatin initiators based on time since prior MI can be applied to extrapolate rates for normostatin and fantastistatin had patients with longer intervals between MI and treatment initiation been eligible to participate in the trial (Table 2). An average estimate of efficacy in

a defined target population can then be estimated by reweighting the observed and extrapolated rates of events to match the distribution of age and/or interval between MI and treatment initiation in the target population.

Discrete event simulation

Discrete event simulation (DES) is a method for modeling disease pathways and outcomes over time as a function of treatment and patient-level covariates. DES keeps track of patient-level covariates and accounts for the changes in patients' risk factors over time. Event rates, absolute risks, and treatment effect sizes can then be estimated based on predefined functional forms that describe the relationships between risk factors and outcomes.^{16,26–28} For example, the fact that the risk of MACEs increases as a patient ages or accumulates comorbidities can be incorporated as parameters when estimating individual risks for each patient as they transition through different pathways and health states in the simulation. Thus, DES can extrapolate over extended time spans, as patients accumulate risk-modifying characteristics, such as older age or greater comorbidity burden. Similar to reweighting, DES can generalize results to a target population with a different distribution of characteristics than the RCT population and, like cross-design synthesis, DES can also extrapolate RCT evidence to populations excluded from trials (Figure 1c). DES has also explicitly accounts for transitions in health status over time when estimating risk. For example, when estimating risk of stroke over 5 years, DES can incorporate changes in other stroke risk factors within the 5-year interval, and allow this to inform the estimated 5-year risk of stroke. Evaluation of extrapolated treatment effects in a defined target population can be achieved by making predictions for a real or simulated population with the relevant distribution of characteristics.

The steps for DES include:

1. Develop and validate multivariable outcome prediction models that describe the relationship between patient covariates and outcomes for each exposure group of interest using trial data directly and/or external information for factors that were exclusion criteria in the trial (e.g., from published literature or evaluation of observational healthcare data).
2. Design a DES model based on possible pathways through different health states that patients could experience over time. Use the outcome prediction models from step 1 to define probabilities of outcomes based on patient characteristics as those characteristics for each individual change over time.
3. Simulate a cohort of patients with covariate distributions that match the RCT population.
4. Validate the DES by comparing simulation-estimated event rates and effect measures with those from the RCT.
5. Use DES to simulate a cohort of patients with covariate distributions that match the target population in routine care. Marginal distributions for covariates can be obtained from the literature or joint distributions obtained by accessing a large healthcare data source and creating the relevant study population.

6. Run the DES to obtain predicted absolute event rates and effect measures in the target population. Uncertainty of estimates from each step in the DES is carried forward by simulating the next step at each transition based on model point estimates and standard errors.

Outcome prediction models can be obtained from the literature, developed by fitting regression models and using individual-level data for trial participants or in observational data. Newly developed models using individual-level data should be validated to evaluate how well they predict outcomes in out-of-sample data. A recent study used published outcome models from the Randomized Evaluation of Long-Term Anticoagulation Therapy trial, and baseline characteristics from two previously published observational studies that compared benefits and risks of dabigatran to warfarin in patients with atrial fibrillation to develop a DES model that was able to replicate the rates of ischemic stroke and major bleeding observed in the Randomized Evaluation of Long-Term Anticoagulation Therapy trial. This model was then used to predict what the results would have been had the trial been conducted in populations similar to those observed in the target populations observed in routine care.¹⁶

In our fictional case study, investigators had access to individual-level data from both RCT and observational data and were able to develop and validate outcome models to quantify the effect of age, time from prior MI to treatment initiation, and exposure within trial data as well as a combination of trial data and observational data. One of the many licensed or open source software tools currently available could then be used to develop a DES model that simulates processes over time for the target population of interest in routine care. This model could then be used to generate extrapolated estimates for older patients and patients with longer lags between MI and treatment initiation that were excluded from the pivotal trial.

DISCUSSION

After a drug has gained regulatory approval and entered the market, it can take a few years for there to be sufficient use of the product to conduct postapproval evaluation of effectiveness and safety in a routine care population that often includes patients who are quite different in terms of disease state, comorbidity, and treatment path from those who were selected for trials. Using the methods outlined in this review, earlier evidence to better understand the expected efficacy of the drug in an expanded routine care population can be generated nearly concurrently with conclusion of a pivotal trial. However, it should be recognized that generalized or extrapolated evidence based on trial efficacy may not reflect effectiveness in real-world practice due to difference in adherence, outcome measurement, pathophysiological differences between populations, or limitations of the extrapolation methods.

The methods that we reviewed share some common strengths, such as the ability to provide estimated measures of occurrence (such as rates) and measures of effect (such as risk differences and risk ratios) for populations different from the enrolled trial participants. Each of these methods can provide estimated measures of occurrence and effect, had the characteristics of the trial participants been distributed differently. The methods also have

some shared limitations. For instance, they all require that relevant out-comes and risk factors are measured or measurable and that the algorithms used to define these conditions capture approximately the same clinical concepts in trial data and observational data. Additionally, when the target population has characteristics that are under-represented in the trial, it can result in a few trial participants having a great deal of influence on weights or extrapolation algorithms. These methods also cannot overcome certain limitations of RCTs, such as the inability to detect rare adverse events in small trial populations or those with delayed onset.

Each method also has unique strengths and limitations (Table 3). Reweighting methods have the advantage of not requiring fitting and validating complex outcome models. Instead, modeling is focused on predicting how baseline characteristics relate to probability of being eligible for inclusion in the relevant trial vs. treatment in routine care. Particularly for rare outcomes, or other outcomes where it is difficult to develop a good outcome prediction model, this can be a strong advantage. However, the PS model to estimate weights can be complex. Additionally, reweighting methods cannot extrapolate estimates to populations with characteristics that were excluded from the trial. In contrast, both cross-design synthesis and DES can extrapolate to populations that were excluded from trials.

Extrapolation in cross-design synthesis and DES approaches assumes that there is no confounding or modification of the effect measure of interest by inclusion criteria for the trial (or unmeasured correlates of those criteria) and that observed trends in relationships between patient characteristics and the outcome continue unchanged in areas of extrapolation.¹⁵ The worse the performance of outcome prediction models in the types of patients that are observed (in terms of calibration and discrimination), the lower the confidence in predictions for types of patients that were not observed. Depending on how the models are fit, or where information for models and extrapolation comes from, both cross-design synthesis and DES extrapolation could potentially mix estimates of efficacy (from trials) and effectiveness (from observational data). Trends in treatment efficacy across age groups (or other characteristics) measured in trials may or may not be similar to trends for effectiveness in the real-world setting. For example, patient characteristics related to nonadherence to therapy may be related to age or lag between MI and treatment initiation. A participant that is closely monitored in a trial may be more adherent than one would as part of routine care.

Although these methods are not a substitute for less restrictive pre-approval RCTs or rigorous observational studies when sufficient data are available in the postapproval setting, they can help to fill the evidence gap that exists in the early marketing period. Acquiring access to observational data and creating relevant study populations for one or more comparator drugs can be accomplished while the trial is ongoing. Analysis plans and code can be written for each of the reviewed methods before a trial is completed so that when the final, analytic dataset for the trial is ready, the code can be executed in parallel with the main trial analysis. However, early evidence using real-world data and methods for generalizing or extrapolating evidence should be reported with clear explanation of assumptions and limitations, especially when used to support regulatory and health technology assessment decisions.

Acknowledgments

FUNDING

This review was funded by Bayer.

References

1. Herrera AP, Snipes SA, King DW, Torres-Vigil I, Goldberg DS & Weinberg AD Disparate inclusion of older adults in clinical trials: priorities and opportunities for policy and practice change. *Am. J. Public Health* 100(suppl 1), S105–S112 (2010). [PubMed: 20147682]
2. Avorn J Including elderly people in clinical trials: better information could improve the effectiveness and safety of drug use. *BMJ* 315, 1033 (1997). [PubMed: 9366725]
3. Bourgeois FT, Orenstein L, Ballakur S, Mandl KD & Ioannidis JPA Exclusion of elderly people from randomized clinical trials of drugs for ischemic heart disease. *J. Am. Geriatr. Soc.* 65, 2354–2361 (2017). [PubMed: 28306144]
4. Kautzky-Willer A, Harreiter J & Pacini G Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. *Endocr. Rev.* 37, 278–316 (2016). [PubMed: 27159875]
5. Phelan AL, Kunselman AR, Chuang CH, Raja-Khan NT & Legro RS Exclusion of women of childbearing potential in clinical trials of type 2 diabetes medications: a review of protocol-based barriers to enrollment. *Diabetes Care* 39, 1004–1009 (2016). [PubMed: 27208329]
6. Charytan D & Kuntz RE The exclusion of patients with chronic kidney disease from clinical trials in coronary artery disease. *Kidney Int.* 70, 2021–2030 (2006). [PubMed: 17051142]
7. Rothman KJ, Greenland S & Lash TL *Modern Epidemiology* 3rd edn (Lippincott, Williams & Wilkins, Philadelphia, PA, 2008).
8. Schneeweiss S & Avorn J A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J. Clin. Epidemiol.* 58, 323–337 (2005). [PubMed: 15862718]
9. Eichler HG et al. Adaptive licensing: taking the next step in the evolution of drug approval. *Clin. Pharmacol. Ther.* 91, 426–437 (2012). [PubMed: 22336591]
10. Gagne JJ et al. Active safety monitoring of newly marketed medications in a distributed data network: application of a semi-automated monitoring system. *Clin. Pharmacol. Ther.* 92, 80–86 (2012). [PubMed: 22588606]
11. Nallamothu BK, Hayward RA & Bates ER Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. *Circulation* 118, 1294–1303 (2008). [PubMed: 18794402]
12. Eichler H-G et al. Bridging the efficacy-effectiveness gap: a regulators perspective on addressing variability of drug response. *Nat. Rev. Drug Discovery* 10, 495 (2011). [PubMed: 21720406]
13. Kent DM, Rothwell PM, Ioannidis JPA, Altman DG & Hayward RA Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 11, 85 (2010). [PubMed: 20704705]
14. Stuart EA, Cole SR, Bradshaw CP & Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *J. R. Stat. Soc. Ser. A Stat. Soc.* 174, 369–386 (2001).
15. Kaizar EE Estimating treatment effect via simple cross design synthesis. *Stat. Med.* 30, 2986–3009 (2011). [PubMed: 21898521]
16. Najafzadeh M, Schneeweiss S, Choudhry NK, Wang SV & Gagne JJ Simulation for predicting effectiveness and safety of new cardiovascular drugs in routine care populations. *Clin. Pharmacol. Ther.* 10.1002/cpt.1045.
17. Cole SR & Stuart EA Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am. J. Epidemiol.* 172, 107–115 (2010). [PubMed: 20547574]
18. Stuart EA, Bradshaw CP & Leaf PJ Assessing the generalizability of randomized trial results to target populations. *Prev. Sci.* 16, 475–485 (2015). [PubMed: 25307417]

19. Toh S, Gagne JJ, Rassen JA, Fireman BH, Kulldorff M & Brown JS Confounding adjustment in comparative effectiveness research conducted within distributed research networks. *Med. Care* 51(8 suppl 3), S4–S10 (2013).
20. Sturmer T, Wyss R, Glynn RJ & Brookhart MA Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs. *J. Intern. Med.* 275, 570–580 (2014). [PubMed: 24520806]
21. Guestrin C, Bodik P, Thibaux R, Paskin M & Madden S Distributed regression: an efficient framework for modeling sensor network data. Paper presented at: Information Processing in Sensor Networks, 2004. IPSN 2004 Third International Symposium on 2004.
22. Verde PE & Ohmann C Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. *Res. Synth. Methods* 6, 45–62 (2015). [PubMed: 26035469]
23. Henderson NC, Varadhan R & Weiss CO Cross-design synthesis for extending the applicability of trial evidence when treatment effect is heterogenous: part II. Application and external validation. *Commun. Stat. Case Stud. Data Anal. Appl.* 3, 7–20 (2017).
24. Peters JL, Rushton L, Sutton AJ, Jones DR, Abrams KR & Muggleston MA Bayesian methods for the cross-design synthesis of epidemiological and toxicological evidence. *J. R. Stat. Soc. Ser. C Appl. Stat.* 54, 159–172 (2005).
25. Prevost TC, Abrams KR & Jones DR Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Stat. Med.* 19, 3359–3376 (2000). [PubMed: 11122501]
26. Dahabreh IJ, Hayward R & Kent DM Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int. J. Epidemiol.* 45, 2184–2193 (2016). [PubMed: 27864403]
27. Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG & Hayward RA Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *Int. J. Epidemiol.* 45, 2075–2088 (2016). [PubMed: 27375287]
28. Burke JF, Hayward RA, Nelson JP & Kent DM Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. *Circ. Cardiovasc. Qual. Outcomes* 7, 163–169 (2014). [PubMed: 24425710]

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Many patient populations that are major consumers of drugs are underrepresented in randomized trials, however evidence from trials are the basis for regulatory, clinical and other decisions that have important implications for patients.

WHAT QUESTION DID THIS STUDY ADDRESS?

We review three methods for extrapolating evidence from trial participants to target populations treated in routine care.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

These methods could help fill the evidence gap for underrepresented populations in the early post-approval setting.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

Methods to generalize the baseline characteristics of the RCT population to patients treated in routine care can facilitate generation of early evidence for the effectiveness and safety of treatments in excluded populations.

Figure 1.

Distribution of age and time from prior myocardial infarction (MI) to treatment initiation in randomized controlled trial (RCT) participants and target population in routine care. **(a)** Observed distribution in RCT participants and target population in routine care. **(b)** RCT participant data reweighted to distribution observed for routine care population in area of overlap (within dashed box). **(c)** Extrapolation from RCT participants to target routine care population.

Figure 2.

Imagined smooth locally weighted scatterplot smoothing curves fit through incidence rates after initiation of statins already on the market or fantastistatin stratified by age. Dotted lines represent extrapolation with cross-design synthesis. Dashed lines represent linear extrapolation. MACE, major adverse cardiac event.

Table 1
Imagined rate of MACE per 1,000 person years following treatment initiation stratified by age

Age	Observed data			Extrapolated data		
	A. Observational data	B. Randomized trial data	C. Randomized trial data	D. Cross-design synthesis	E. Cross-design synthesis	
	Normostatin	Normostatin	Fantastistatin	Normostatin	Fantastistatin	
40-44	8.9	7.1	6.2	7.1	6.2	
45-50	8.9	7.1	6.2	7.1	6.2	
55-59	9.1	7.2	6.3	7.2	6.3	
60-64	9.3	7.4	6.4	7.4	6.4	
65-69	9.7	1.05		7.8	1.05 6.8	1.05
70-74	10.7	1.10		8.5	1.10 7.4	1.10
75-79	12.3	1.15		9.8	1.15 8.5	1.15
80-84	14.8	1.20		11.8	1.20 10.2	1.20
85-89	18.4	1.25		14.7	1.25 12.8	1.25
90+	23.1	1.25		18.4	1.25 16.0	1.25

MACE, major adverse cardiac event.
= Relative change in estimated rate of MACE for sequential strata (moving from top to bottom).

Table 2

Imagined rate of MACE per 1,000 person years following treatment initiation stratified by time since prior MI

	Observed data			Extrapolated data	
	A. Observational data	B. Randomized trial data	C. Randomized trial data	D. Cross-design synthesis	E. Cross-design synthesis
	Normostatin	Normostatin	Fantastistatin	Normostatin	Fantastistatin
Time since prior MI					
0–3 months	8.4	7.5	6.4	7.5	6.4
4–6 months	8.0	0.95		7.1	6.1
7–9 months	7.6	0.95		6.8	5.8
10–12 months	6.8	0.90		6.1	5.2
				0.90	0.90

MACE, major adverse cardiac event; MI, myocardial infarction.

= Relative change in estimated rate of MACE for sequential strata (moving from top to bottom).

Table 3

Comparison of methods to generalize or extrapolate trial estimates to patients in routine care

	Re-weighting	CDS	DES
Strengths			
Can provide estimates for a defined population based on numerous patient characteristics	Yes	Yes	Yes
Can provide estimates for absolute or relative effects	Yes	Yes	Yes
Accounts for joint distribution of patient characteristics if individual level data accessible	Yes	Yes	Yes
Simpler analyses can be conducted with public summary information	Yes	Yes	Yes
Can extrapolate beyond characteristics observed in trial data	n/a	Yes	Yes
Models time varying risk as patients transition between health states	No	No	Yes
Limitations			
Major risk factors for the outcome must be measured in the trial, and measurable in observational data	Yes	Yes	Yes
Algorithms to capture risk factors in observational data capture the same clinical concepts as the trial	Yes	Yes	Yes
Must develop (and validate) good outcome models that incorporate relevant risk factors	N/A	Yes	Yes
Must develop good propensity score model that accounts for important confounders	Yes	N/A	N/A
May mix estimates of efficacy and effectiveness	N/A	?	?
Assumptions			
Models are correctly specified (negligible unmeasured confounding)	Yes	Yes	Yes
Negligible confounding or effect modification by inclusion criteria for the trial (or unmeasured correlates)	N/A	Yes	Yes
Trend in rate of outcomes continues unchanged in area of extrapolation	N/A	Yes	Yes
Trends in rate of outcome given patient characteristic(s) of interest are not modified by	N/A	?	?

?, depends on situation; CDS, cross-design synthesis; DES, discrete event simulation; N/A, not applicable.