

## **STRATEGIES TO MEET THE NEED FOR LONG-TERM DATA**

John Chalmers<sup>1</sup>, Mark Woodward<sup>1-3</sup>, Claudio Borghi<sup>4</sup>, Athanasios Manolis<sup>5</sup>, Giuseppe Mancia<sup>6,7</sup>

<sup>1</sup> The George Institute for Global Health, University of Sydney, Australia

<sup>2</sup> The George Institute for Global Health, University of Oxford, UK

<sup>3</sup> Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA

<sup>4</sup> Department of Medical and Surgical Sciences, University of Bologna, Italy

<sup>5</sup> Department of Cardiology, Asklepeion Hospital, Athens, Greece

<sup>6</sup> University of Milano-Bicocca, Milan, Italy

<sup>7</sup> Centre of Epidemiology & Clinical Trials, IRCSS Istituto Auxologico Italiano, Milan, Italy

Short Title:                   **Meeting the Need for Long-term Data**

Corresponding Author  
Prof John Chalmers  
Senior Director  
The George Institute for Global Health  
PO Box M201  
Missenden Road   NSW   2050  
Australia  
Email: [chalmers@georgeinstitute.org.au](mailto:chalmers@georgeinstitute.org.au)  
Phone: +61 2 9993 4587 (work)  
          +61 2 405 196 000 (mobile)  
Fax:    +61 2 9993 4588

Word Count: 4010

## **ABSTRACT**

Chronic diseases afflict patients for many years, often to the end of life, and there is increasing need for estimating lifelong risk and for evaluating the effects of treatment in the long-term. Yet recommendations for lifelong treatment are most frequently based on findings from randomised clinical trials lasting only a few years. There is therefore a clear need for much longer-term data, and here we present the advantages and disadvantages of many strategies, including the use of long-term post-trial follow-up, of long-term prospective cohort studies, registry databases, and of administrative data bases. We also emphasise the need for long-term cost-effectiveness studies. One of the most promising strategies comes from linkage of data gathered through the ever-expanding pool of administrative data bases worldwide with data from other sources, including randomised trials and the many forms of observational study.

Keywords: long-term follow-up; randomised controlled trial; registry; database; cohort study; administrative database; cost effectiveness study

Chronic diseases affect patients for many years, often attended by exacerbations and remissions, and most guidelines recommend such patients should undergo lifelong treatment. This is clearly so for hypertension, diabetes and dyslipidaemia which are characterized by a long life expectancy. It also applies to patients with coronary artery, cerebrovascular and renal diseases whose survival now spans over many years. However, the evidence base used to justify and to guide treatment, particularly pharmacological treatment, is heavily dependent on randomised controlled trials (RCTs), commonly providing 3-6 years of active treatment and follow-up. This bias towards dependence on RCTs has increased in recent years with greater insistence that guideline recommendations for treatment should be based on the highest levels of evidence, and thus on data from RCTs and on meta-analyses of such trials, as for example in the latest recommendations from the members appointed to the 8<sup>th</sup> Joint National Committee (“JNC8”) (1-4). Cost effectiveness data are also essential in formulating recommendations for lifelong treatment, but at present these are subject to the same limitations as those affecting clinical risk/benefit data, including undue dependence on relatively short-term RCTs.

There are many other forms of evidence that should be considered in guiding life-long treatment, all of them observational in nature. These include post-trial follow-up of participants in RCTs, after the period of randomised treatment comparison has ceased. Other valuable sources of evidence include data from well-conducted national or regional registries, the use of cohort studies with longitudinal follow-up, and - increasingly possible - the use of data from large administrative data bases that are collected by health authorities around the world, albeit with substantial variation in the range and quality of the data that are accumulated. The purpose of this review is to discuss the use of these strategies and to comment on their advantages and disadvantages.

### *Long-term post-trial follow-up*

One very useful strategy, used over the past two decades, has been the long-term follow-up of participants in RCTs for lengthy periods after randomised treatment has ceased. This was initiated in studies of more intensive glucose control in patients with diabetes, both type 1 and type 2 diabetes, motivated by the argument that the 5-10 year period of active comparison might be insufficient to detect beneficial effects of superior metabolic control, which might take many more years to become manifest (5,6). These initiatives were rewarded by the detection of long-term reductions in major cardiovascular complications that were not evident by the end of the period of randomised treatment (7-9). This has been reinforced by recent reports from the Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation Observational Study (ADVANCE-ON) of persistent long term benefits of more intensive glucose control for end-stage kidney disease (ESKD) and from the Veteran's Affairs Diabetes Trial (VADT) study of long-term benefits of tighter glucose control for major vascular events. (10,11) ADVANCE-ON has also reported long-term, post-trial mortality benefits following blood pressure lowering with the combination of perindopril and indapamide (10). One of the advantages of long-term post-trial follow-up is that it enables investigators to confirm or refute suggestions from the later phase of cumulative event curves obtained during the randomised treatment phase, that the curves for the active treatment and the control groups are either beginning to diverge or starting to converge. This is well illustrated by the ADVANCE-ON study, which clearly showed that there was no long-term benefit for major macrovascular events over a period of 10 years following more intensive glucose lowering (Figure 1, right panel), even though there was a suggestion of emerging benefit during the last year of randomised treatment (Figure 1, left panel) (10,12).

Over the past two decades, many trials of blood pressure lowering drugs, as well as trials of statins, have conducted long-term post-trial follow-up for between 1 and 10 years and yielded much valuable information. They have demonstrated that some of the benefits of randomised treatment do persist beyond the period of active comparison, although these benefits progressively attenuate or dissipate over time (10,13-22). While this attenuation may reflect the lack of a strong “legacy effect”, it is also testimony to the efficacy of active treatment during the randomised, in-trial period, and to the importance of maintaining active blood pressure lowering in the long-term. These post-trial data help extend our experience and available evidence regarding the treatment of chronic cardiovascular diseases to around one decade or more, but this still falls well short of the evidence base needed to justify and guide lifelong treatment. There are also some inherent disadvantages in post-trial follow-up, including the problems of loss to follow-up and of drop-outs.

The great advantage of long-term post-trial follow-up, for the long-term study of possible risks and benefits of earlier periods of treatment, lies in the reduction of biases through maintenance of the benefits of randomisation. This requires maintenance of the ‘intention to treat’ principle which preserves the close matching of active and control treatment groups, and this, in turn, requires the analysis of the cumulative incidence of events or outcomes, from the very beginning of randomised treatment through to the end of the period of post-trial follow-up. Analysis of data from the period of post-trial follow-up alone loses the advantage of comparing evenly balanced groups of participants derived through randomisation. Instead, it compares disparate groups that have had different experiences during randomised treatment, with substantial confounding likely to be introduced through differences in event rates, including mortality rates, between the groups allocated randomly to different treatments at baseline. One issue that may arise is that when a clinical trial terminates, or is prematurely

stopped due to net advantage of one treatment over another, participants may be transferred to the more advantageous treatment on ethical grounds, thus losing the advantage of “intention to treat” analysis, and requiring or even defying, more complex analytic procedures to dissect out the real long-term benefits of the original active treatment.

There are, however, some clear limitations and disadvantages in predominant reliance on data from participants in RCTs. First and foremost, most investigators agree that there are inherent biases in the selection of participants in clinical trials, which may limit the generalizability and applicability of trial results to broader populations at community level. This is often in the direction of higher risk groups, due to selection of more vulnerable groups, which are chosen to increase event rates and thus study power or to reduce participant numbers and thus cost. However, it can work in the opposite direction since ethical and logistic considerations can influence selection towards fitter and healthier patients. This is particularly true in older populations, such as in the Hypertension in the Very Elderly Trial (HYVET) in patients over the age of 80 years, which deliberately excluded very frail elderly patients who might be at greater risk of the adverse effects of more stringent blood pressure lowering (23). It also applies to many trials of treatment for heart failure which are carried out in patients aged in the early 60s, in contrast with the much older and physically compromised heart failure patients who are usually seen and treated in clinical practice. Another disadvantage of RCTs is the very great cost of large prospective outcome trials, often beyond the funding capacity of most national research agencies, and increasingly challenging for many pharmaceutical companies. The need to fund the additional cost of many years of post-trial follow-up will often prove to be the straw that breaks the camel’s back! Furthermore, the costs must clearly limit the length of follow-up. In summary, prospective RCTs clearly provide the gold standard in recommendations for treatment, but they have definite disadvantages and their very great cost will clearly limit the practicality of using them directly, or through post-trial

follow-up, as the chief source of truly long-term evidence for treatment of cardiovascular conditions.

### ***Long term prospective cohort studies***

Cohort studies are well-recognised as the most reliable type of observational study. Whilst not having the benefit of randomization to guarantee, with large numbers, lack of bias by indication, they provide a valid source of information on the effects of medications and lifestyle factors when analyses are suitably controlled for confounding. Provided that membership of the study population is unlikely to modify the association between exposure and outcome, even special populations, such as employees in a specific industry, can provide meaningful results. Non-random samples are, however, not able to accurately estimate absolute risks. This is important because, increasingly, the take up of invitations to join cohort studies has declined with response rates of 20% or less being reported (24,25). This makes claims for “general population” coverage of such cohort studies debatable, and potentially misleading.

Although such cohort studies are generally started with a fixed time horizon, usually decided according to the resources initially available, they are often extended so as to give long-term follow-up appropriate to analysing the effects of long-term exposure. For example, The Scottish Heart Health Study originally reported on events and associations after around 8 years (26), but the most recent report from this study uses 22 years of follow-up data (27). However, that study only used exposure data from baseline, so that it is unreasonable to assume that these data represent long-term continuous exposures.

Other cohort studies have a longitudinal form wherein surviving subjects are resurveyed to update their data on medication use, lifestyles etc, at intervals. The Framingham Heart Study (28) is a classic example of this, where subjects have been resurveyed every second year

post-baseline. The Atherosclerosis Risk in Communities (29) and Electricity Generating Authority of Thailand (EGAT) (30) studies are further examples. These studies do allow long-term exposures to be evaluated directly. Note, however, that this is only true if the exposure of interest was recognised when the study started – or, at least, was apparent as of potential interest many years ago. Birth cohort studies, such as the Avon Longitudinal Study of Parents and Children (31), are perhaps the ultimate longitudinal studies, although they clearly have to be very mature to enable the effects of medications typically given in middle age, or later, to be studied. Framingham and EGAT have the further feature of recruiting new cohorts from the same population, over time, which can allow analyses of long-term continuous effects, for example to drugs developed 10-20 years ago, in relatively young people.

A modern trend is to initiate Biobanks. Often these are hospital-based, but general population Biobanks, such as the China Kadoori Biobank (32) and the UK Biobank (33), are beginning to provide excellent research data. Except for the fact that such studies typically have a wider range of exposures, they are essentially cohort studies, with a baseline survey followed by monitoring of participants, often using routine data sources. Long-term data from such studies are awaited.

### ***Administrative Databases***

Electronically stored routinely-collected data are becoming increasingly used in medical research. Such data have been kept by health services for very many years, yet rarely used for research, as opposed to administrative, purposes. Although the earliest use of such data for research purposes may be attributed to the US Veterans Association, the methodology was developed by the seminal Oxford Record Linkage Study (34). In the main, these data are of two types: episode-based data, such as the Clinical Practice Research Datalink (CPRD) in the



UK (35), and billing-based, such as from the several health-care systems in the USA, for example, from the Kaiser Permanente system (36). In essence, these are similar to cohort studies, albeit developed for research purposes retrospectively. For example, Emdin *et al* (37) used CPRD to identify 4.1 million patients who had at least one blood pressure measurement taken by their general practitioner (GP) between 2000 and 2012. These patients were followed retrospectively through to 2012 using linked digital record systems covering hospital visits and death registrations, as well as future GP consultations. CPRD is a rich data source because of its myriad linkages; most other routine data are more restricted, for example to hospitalizations. Linkages are essential to get a comprehensive coverage of events.

Many long-term cohort studies depend on linkage to routine data, collected by administrative data systems, to capture pharmaceutical prescriptions as well as major events, including hospitalisation and death, further emphasising the similarity of the two approaches to research (Figure 2). However, several key differences are apparent between these two strategies. First, pure use of electronic health systems requires no baseline survey. Instead, patients enter the study when they first encounter the health (or other) system which uses the routine data – when they first visit their doctor in the case of CPRD. This clearly has enormous savings in cost. Second, there is no need to wait many years to get long-term data; provided the database has been running enough time one can start analyses straightaway. Third, unlike any bespoke research data, there is no appreciable cost penalty for increasing the sample size. Fourth, longitudinal observational studies (and RCTs) have limited capacity to address the problem of low adherence to drug treatment of chronic diseases in everyday clinical practice, because recruitment to such studies, and hence awareness of being under observation, alters patients' behaviour. Where drugs are free of charge (and thus self-purchase is negligible) this may be addressed using administrative data-bases, which can

quantify adherence to a treatment regimen through the proportion of the follow-up time covered by prescription (38). Recently, this strategy has documented that in the general population, antihypertensive, lipid lowering and anti-diabetic treatments are all characterized by a very low adherence level, with a substantial adverse impact on the risk of hospitalization for cardiovascular disease (39,40). Each of these four features highlights the advantages of strategies that use information from administrative databases. However, the following three differences between routinely collected data and cohort data point to the limitations of administrative databases. First, the clinical information gathered in administrative data bases is often limited and is of variable quality (38). Second, missing values are generally much more frequent in administrative data bases than in cohort study data, often because a patient only gets the tests that their doctor requests at any point in time. A prime example is where QRESEARCH, a UK database similar to CPRD, was used to define a cardiovascular risk score (41). Due to a large amount of non-randomly missing data on lipids in QRESEARCH, and an inadequate analytical procedure to deal with this, the emergent score substantially underplayed the importance of the total to high-density lipoprotein cholesterol ratio in the subsequent risk prediction algorithm (42). The third, and possibly the most important difference between the two types of data, is a feature which may make routine data unsuitable: the problem of indication bias. As noted, a standard feature of most routine data used in medical research is that many participants enter the study when something happens to them, typically an illness. This may make them more likely to receive a battery of tests, meaning that the baseline values of study variables are not available from even an unbiased sample, let alone a representative one of the study population. The sicker the patient, the more likely it is that the doctor will administer an (in some way) more extreme treatment (or, perhaps, a less conservative regimen), so that comparisons of effects of medications may well be biased. The problem of indication bias is lessened, but not removed, when it is mandatory

that everyone should be registered with the administrative system in question. Another issue that may complicate long-term follow-up, lies in the use of drugs with multiple actions, such as blood pressure lowering drugs that have additional properties favouring their use in particular cardiovascular conditions, such as coronary heart disease, where beta-blockers may have advantages, or heart failure where ACE inhibitors or diuretics may confer additional benefits. Subtle biases in estimates of interaction effects are also likely. Hence the long-term effects of treatment may well be analysed with greater statistical precision, but less practical reliability (i.e. sampling error is reduced but bias is increased). Although increasingly addressed by using propensity scoring (43), there is much uncertainty as to whether this can actually resolve the problem, given that the limited clinical information offered by administrative databases restricts the possibility of matching groups on a sufficiently large number of clinical variables (38). Other possibilities such as restricting the investigated cohort to patients experiencing the nominated study outcome (the case-only design), sensitivity analysis and new analytic techniques, such as instrumental variable methods have also been applied to administrative databases to minimise confounding, with variable results (38).

Concerted action is needed, involving cooperation between health researchers and health administrators, to increase the quantity and quality of the data that governments like to collect for financial and logistic purposes. This is essential to make them more comprehensive so as to be suitable for clinical research, and also to make it more accessible and available for legitimate research studies. A fine balance must be struck between the need for data confidentiality, limits to accessibility and the needs of medical research – and we are still learning how to achieve this (44, 45). While this should happen at national level, there are also advantages in regional systems that make such data available for health systems research, preferably as part of an overall national database.

### ***Registry Databases***

By themselves, registries will not yield the long-term data that are needed. However, in combination with randomised trials, cohort studies or with linkage to routine collection systems, such as electronic health records, they can provide a very valuable platform with the potential for effective long-term follow-up. (46,47). To be effective, registries must be of high quality and very few around the world at present meet the standards required. Some of the best in the cardiovascular field appear to be those in Scandinavian and in the UK at present, although there have long been many excellent registries for cancer and for reportable (or “notifiable”) infectious diseases, across the world.

The use of a high-quality registry as the initial platform for an RCT can markedly reduce the complexity, expense, time to recruitment and also improve the representativeness of the trial, especially when it can be linked to further data bases with information regarding deaths and other major outcomes. One such recent example was the “Thrombus Aspiration during ST-Segment Elevation Myocardial Infarction in Scandinavia” (TASTE) trial, which was based on the national comprehensive Swedish Coronary Angiography and Angioplasty Registry (SCAAR) with endpoints evaluated through further national Swedish registries, including mortality data from the national population registry, and major clinical endpoints from the internet based Swedish Web System for Enhancement and Development of Evidence-based Care in Heart Disease Evaluated According to Recommended Therapies (SWEDEHEART) (47,48). Combined use of such resources, not only facilitates the conduct and reduces the cost of a major RCT, it also makes long-term follow-up or participants feasible, even to their eventual and inevitable death.

Furthermore, just as judicious use of high quality registries can be combined with randomised trials, so can it be combined with cohort studies and with linkage to “Big Data” such as the CPRD in the UK, to provide very effective and more affordable means of long-term follow-up. The West of Scotland Coronary Prevention Study (WOSCOPS) provides a prime example that such combination works well and that routinely collected data, in this instance through the Scottish Health Informatics Programme, can provide high quality and valuable information (49). Considerable challenges remain with such integration, particularly for use with RCTs. This includes issues with patient-level (individual) randomisation, with the comfort level of the health professionals actually caring for the patient, with privacy and informed consent of participants, and with ‘blinding’ of investigators. (50).

### *Cost effectiveness studies*

Cost effectiveness data are also critical for meaningful recommendations, especially in low resource settings, and indeed across the world in times of escalating health costs and difficult economic circumstances. However at present, cost effectiveness data are heavily dependent on short-term studies, especially clinical trials, with many estimates of long-term cost-benefit based on complex modelling involving many assumptions (51). Thus the use of long-term costs data suffers from the same limitations as those applying to clinical data on efficacy and adverse effects, and is equally in need of effective solutions for the acquisition of long-term data, that can be used for determining cost-effectiveness over a life-time. One of the great advantages of long-term data would be that it would allow the adverse consequences of treatment to be more appropriately measured including unexpected adverse effects that only become apparent in the long-term such as, for example, development of type 2 diabetes with use of antihypertensive drugs or statins (52,53). In this way, the true costs, both monetary

and clinical, could be estimated and the cost benefit ratio made more meaningful and accurate than when only short-term data are available.

In the first instance it would be helpful to build cost-effectiveness data collection into the planning of RCTs, of post-trial follow-up studies and of prospective cohort studies, a feature that is most often omitted or at best rudimentary. However, this is one area where data linkage can be particularly useful, especially for the collection of long-term data, with linkage to national or regional data bases for both clinical and for financial data. Such databases will most often be the administrative data bases linked to government health schemes, but can also be accessed through a variety of ‘private’ health insurance schemes (54). Administrative databases would also have the advantage that long-term information originates from real life medical practice, where drug-related adverse effects may be more common and severe than those detected in RCTs in which greater physicians’ expertise and closer patients’ follow-up will usually ensure greater patient safety.

Another major issue facing investigators wishing to address these cost-effectiveness considerations is the fact that the data are often split between very different jurisdictions/health funders, including multiple levels of government (national, regional, and local) and a variety of private insurers (55). Furthermore, the data on costs in primary care and costs for hospital admissions often involve very different levels of health care providers and funders. Another limitation will be that many of these data bases will likely be restricted to “reimbursable costs” and will omit out-of-pocket costs, such as over-the-counter medication costs, co-payments, and loss of productivity, including salary or employment that may result from serious illness for both the actual patient and for carers. It seems possible, or even likely, that burgeoning health costs and finite resources will force nations to build more comprehensive and higher quality data bases, gathering both clinical data and financial data, in order to permit more rational decisions on the allocation of resources. The nature and the

value of the data collected from new data bases that might result, will be very strongly influenced by the perspective of the planners—whether this be from the viewpoint of the health system itself , or of the funder (likely the government), or of society as a whole.

### ***Conclusions***

Many chronic diseases afflict patients for many years, often for decades, and commonly until the end of their lives. Yet recommendations for lifelong treatment are most frequently based on findings from randomised clinical trials lasting only a few years. There is therefore a clear need for much longer-term data, and we have presented the advantages and disadvantages of many strategies, including the use of long-term post-trial follow-up, of long-term prospective cohort studies, registry databases, and of administrative data bases. We have also emphasised the need for long-term cost-effectiveness studies. One of the most promising strategies comes from linkage of data gathered through the ever-expanding pool of administrative data bases worldwide with data from other sources, including randomised trials and the many forms of observational study. It will be important for more national governments and health systems to develop well maintained, high quality health care data bases, and also for investigators who are planning clinical trials or observational studies to incorporate linkage with health care data bases, both national and regional, during follow-up. Furthermore, some of the greatest benefits may accrue from incorporation of protocols for long-term follow-up into the initial planning of randomised trials and observational studies.

### **Acknowledgements**

Nil

### **Sources of Funding**

Nil

## **Conflicts of Interest**

JC has received research grants from Servier, administered through the University of Sydney, as Principal Investigator for the ADVANCE trial and ADVANCE-ON Post-trial follow-up, and has also received honoraria from Servier for speaking about these studies at Scientific Meetings, all outside the submitted work. MW reports consultancy work for Amgen. CB reports research grants from Menarini and Novartis, honoraria from Sanofi, Menarini, Amgen, Servier, Boehringer, Novartis, MSD, Sigma-Tau, Recordati and membership of Advisory Boards for Amgen, Sanofi, Menarini, Takeda, Novartis, and MSD, all outside the submitted work. GM reports personal fees from Bayer, Boehringer Ingelheim, Daiichi Sankyo, Medtronic, Novartis, Menarini International, Recordati, Servier, Takeda, all outside the submitted work.



## References

1. James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, Handler J, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *J Am Med Assoc* 2014;311:507-520
2. Mancia G, Fagard R, Narkiewicz K, Redon J, Zanchetti A, Bohm M, et al. 2013 ESH/ESC Guidelines for the management of arterial hypertension: the Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens* 2013; 31:1281-1357
3. Weber MA, Schiffrin EL, White WB, Mann S, Lindholm LH, Kenerson JG, et al. Clinical practice guidelines for the management of hypertension in the community: a statement by the American Society of Hypertension and the International Society of Hypertension. *J Hypertens* 2014; 32:3-15
4. Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Institute of Medicine. Clinical Practice Guidelines We can Trust. Washington DC: *The National Academies Press*; 2011
5. Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 1993; 329:977-986
6. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998; 352:837-853. [Erratum, *Lancet* 1999; 354:602]

7. Nathan DM, Cleary PA, Backlund JY, Genuth SM, Lachin JM, Orchard TJ, et al. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med*. 2005 Dec 22;353:2643-2653.
8. DCCT/EDIC Research Group, de Boer IH, Sun W, Cleary PA, Lachin JM, Molitch ME, Steffes MW, Zinman B. Intensive diabetes therapy and glomerular filtration rate in type 1 diabetes. *N Engl J Med*. 2011;365:2366-2376.
9. Holman R, Paul S, Bethel MA, Matthews D, Neil A. 10-Year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008; 9;359:1577-1589
10. Zoungas S, Chalmers J, Neal B, Billot L, Li Q, Hirakawa Y, et al. Follow-up of blood-pressure lowering and glucose control in type 2 diabetes. *N Engl J Med*. 2014;371:1392-1406
11. Hayward RA, Reaven PD, Emanuele NV, et al. Follow-up of Glycemic Control and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med*. 2015;373:978
12. Patel A, MacMahon S, Chalmers J, Neal B, Billot L, Woodward M, et al. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med* 2008, 358:2560-2572
13. Kostis WJ, Thijs L, Richart T, Kostis JB, Staessen JA. Persistence of mortality reduction after the end of randomised therapy in clinical trials of BP lowering medications. *Hypertension* 2010;56:1060-1068
14. Holman RR, Paul SK, Bethel MA, Neil HAW, Matthews DR. Long-term follow-up after tight control of blood pressure in type 2 diabetes. *N Engl J Med* 2008;359:1565-1576

15. Appel LJ, Wright JT Jr, Green T, Agodoa LY, Astor BC, Bakris GL, et al. Intensive blood-pressure control in hypertensive chronic kidney disease. *N Engl J Med* 2010; 363:918-929.
16. Brouwers FP, Asselbergs FW, Hillege HL, De Boer RA, Gansevoort RT, Van Veldhuisen DJ, et al. Long-term effects of fosinopril and pravastatin on cardiovascular events in subjects with microalbuminuria: Ten years of follow-up of Prevention of Renal and Vascular End-stage Disease Intervention Trial (PREVEND IT). *Am Heart J* 2011;161:1171-1178
17. Beckett N, Peters R, Tuomilehto J, Swift C, Sever P, Potter J, et al. Immediate and late benefits of treating very elderly people with hypertension: results from active treatment extension to Hypertension in the Very Elderly randomised controlled trial. *BMJ* 2011;344:d7541
18. Simes J, Simes RJ, Hunt D, Kirby A, Tonkin A, Keech A, et al. Long-term effectiveness and safety of pravastatin in 9014 patients with coronary heart disease and average cholesterol concentrations: The LIPID trial follow-up. *Lancet* 2002;359:1379-1387
19. Strandberg TE, Pyorala K, Cook TJ, Wilhelmsen L, Faergeman O, Thorgeirsson G, et al. Mortality and incidence of cancer during 10-year follow-up of the Scandinavian Simvastatin Survival Study (4S). *Lancet*. 2004;364:771-777
20. Sever PS, Chang CL, Gupta AK, Whitehouse A, Poulter NR, ASCOT Investigators. The Anglo-Scandinavian Cardiac Outcomes Trial: 11-year mortality follow-up of the lipid-lowering arm in the U.K. *Eur Heart J*. 2011;32:2525-2532
21. Bulbulia R, Bowman L, Wallendszus K, Parish S, Armitage J, Peto R, et al. Effects on 11-year mortality and morbidity of lowering LDL cholesterol with simvastatin for

- about 5 years in 20 536 high-risk individuals: A randomised controlled trial. *Lancet* 2011;378:2013-2020
22. Margolis KL, Davis BR, Baimbridge C, Ciocon JO, Cuyjet AB, Dart RA, et al. Long-term follow-up of moderately hypercholesterolemic hypertensive patients following randomization to pravastatin vs usual care: The antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT-LLT). *J Clin Hypertens* 2013;15:542-554
  23. Beckett N, Peters R, Fletcher A, Staessen JA, Liu L, Dumitrascu D, et al. Treatment of hypertension in patients 80 years or older. *N Engl J Med* 2008. 358: 1887-1898
  24. Ivers RQ, Blows SJ, Stevenson MR, Norton RN, Williamson A, Eisenbruch M, Woodward M, Lam L, Palamara P, Wang J. A cohort study of 20 822 young drivers: the DRIVE study methods and population. *Injury Prev* 2006;12:385-389
  25. Land MA, Webster J, Christoforou A, Praveen D, Jeffery P, Chalmers J, Smith W, Woodward M, Barzi F, Nowson C, Flood V, Neal B. Salt intake assessed by 24 h urinary sodium excretion in a random and opportunistic sample in Australia. *BMJ Open*. 2014;4:e003720.
  26. Tunstall-Pedoe H, Woodward M, Tavendale R, A'Brook R, McCluskey MK. Comparison of the prediction by 27 different factors of coronary heart disease and death in men and women of the Scottish Heart Health Study. *BMJ* 1997; 315:722-729
  27. Tunstall-Pedoe H, Woodward M, Hughes M, Anderson A, Kennedy G, Belch J, Kuulasmaa K. Prime mover or fellow traveller: 25-hydroxy vitamin D's seasonal variation, cardiovascular disease and death in the Scottish Heart Health Extended Cohort (SHHEC). *Int J Epidemiol*. 2015; 44:1602-1612
  28. <https://www.framinghamheartstudy.org/> (accessed September 9, 2015)

29. <https://www2.csc.unc.edu/aric/> (accessed September 9, 2015)
30. Vathesatogkit P, Woodward M, Tanomsup S, Ratanachaiwong W, Vanavanan S, Yamwong S, Sritara P. Cohort profile: the electricity generating authority of Thailand study. *Int J Epidemiol* 2012;41:359-365
31. <http://www.bristol.ac.uk/alspac/> (accessed September 9, 2015)
32. <http://www.ckbiobank.org/site/> (accessed September 9, 2015)
33. <http://www.ukbiobank.ac.uk/> (accessed March 29, 2016)
34. Acheson ED, Evans JG. The Oxford Record Linkage Study: A review of the method with some preliminary results. *Proc R Soc Med.* 1964; 57: 269–274
35. <http://www.cprd.com/intro.asp> (accessed September 9, 2015)
36. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)*. 2014;33:1123-1131.
37. Emdin CA, Anderson SG, Woodward M, Rahimi K. Usual blood pressure and risk of new onset diabetes: evidence from 4.1 million adults and a meta-analysis of prospective studies. *JACC* 2015; 66:1552-1562.
38. Corrao G, Mancia G. Generating evidence from computerized healthcare utilization databases. *Hypertension* 2015;65:490-498
39. G. Corrao, A. Parodi, F. Nicotra, A. Zambon, L. Merlino, G. Cesana, G. Mancia. Better compliance to antihypertensive medications reduces cardiovascular risk. *J Hypertens* 2011;29:610-618
40. Corrao G, Conti V, Merlino L, Catapano A, Mancia G. Results of a retrospective database analysis of adherence to statin therapy and risk of nonfatal ischemic heart disease in daily clinical practice in Italy. *Clinical Therapeutics* 2010;32:300-310.

41. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007;335:136
42. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011;30:377-399
43. Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: from naive enthusiasm to intuitive understanding. *Stat Methods Med Res.* 2012;21:273-293.
44. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, et al. Who owns the data? Open data for healthcare. *Front Public Health.* 2016;4:7
45. Williams G, Pigeot I. Consent and confidentiality in the light of recent demands for data sharing. *Biom J.* 2016 doi:10.1002/bimj.201500044
46. Lauer MS, D'Agostino RB Sr. The randomized registry trial--the next disruptive technology in clinical research? *N Engl J Med* 2013;369:1579-1581
47. Frobert O, Lagerqvist B, Olivecrona G, Omerovic E, Gudnason T, Maeng M, et al. Thrombus aspiration during ST-Segment elevation myocardial infarction. *N Engl J Med* 2013; 369:1587-1597
48. Jernberg T, Attebring MF, Hambræus K et al. The Swedish Web System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies (SWEDEHEART). *Heart* 2010;96:1617-1621
49. Ford I, Murray H, Packard CJ, Shepherd J, Macfarlane PW, Cobbe SM; Long-term follow-up of the West of Scotland Coronary Prevention Study. *N Engl J Med.* 2007;357:1477-1486.
50. Angus DC. Fusing Randomized Trials With Big Data: The Key to Self-learning Health Care Systems? *JAMA* 2015;314:767-768

51. Drummond M, Sculpher M. Common methodological flaws in economic evaluations. *Med Care* 2005;43(7 Suppl):5-14.
52. Elliott WG, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *Lancet* 2007;369:201-207
53. G. Corrao, B Ibrahim, F Nicotra, D Soranna, L Merlino, AL Catapano, E Tragni, M Casula, G Grassi, G Mancia. Statins and the risk of diabetes: evidence from a large population-based cohort study. *Diabetes Care*. 2014;37:2225-2232.
54. Riley GF. Administrative and claims record as sources of health care cost data. *Med Care* 2009;47 (7 Suppl. 1): S51-55
55. Pearson SA, Pesa N, Langton JM, et al. Studies using Australia's Pharmaceutical Benefits Scheme data for pharmacoepidemiological research: a systematic review of the published literature (1987-2013). *Pharmacoepidemiol Drug Saf* 2015;24:447-455

## Figure Legends

**Figure 1.** Cumulative incidence of major macrovascular events according to glucose control treatment assignment during the randomised trial period for ADVANCE (left panel) and during overall follow-up including the randomised and post-trial periods (right panel). The solid line represents intensive glucose control and the dotted line represents standard, guideline-based, control.

**Figure 2.** Basic strategies for obtaining long-term data: (1) mounting a cohort study, ideally with repeated observation of exposures; (2) using administrative data with linkages, for example from general practitioner records to hospital admissions and death certifications; (3) following subjects up after the randomisation period expires in a clinical trial, either through an observational (“cohort”) period or using linked administrative data.



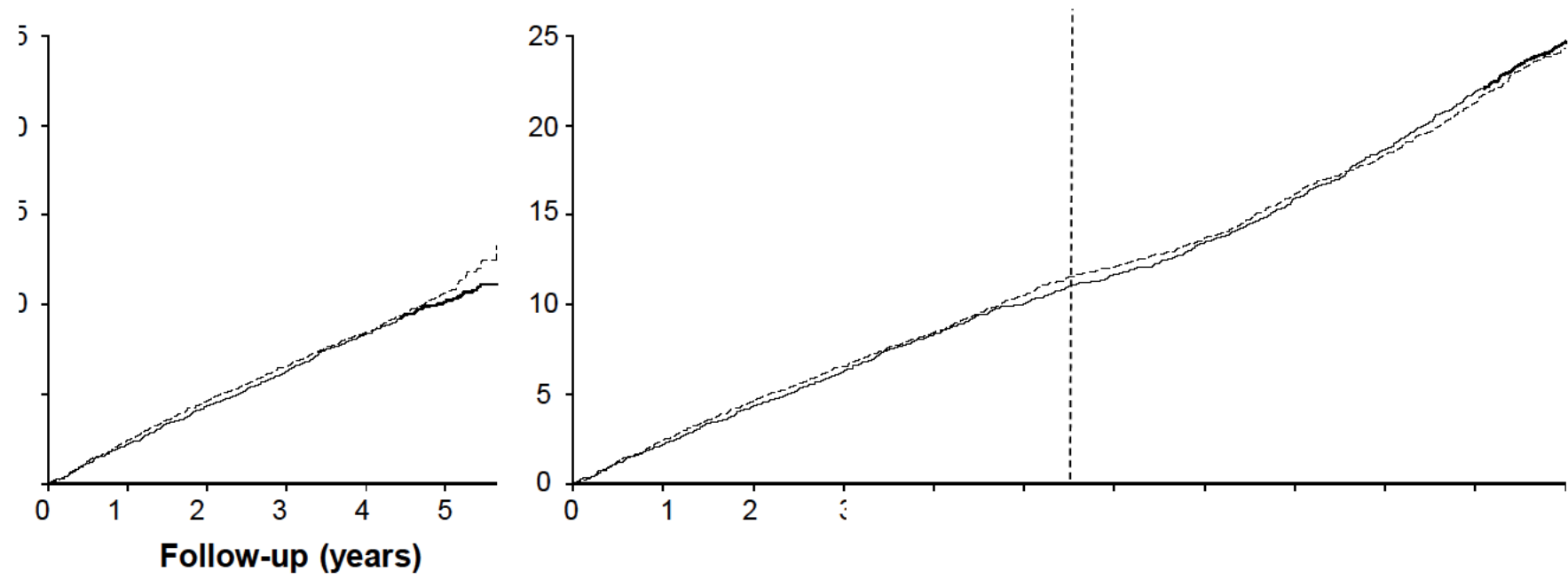


Figure 1.

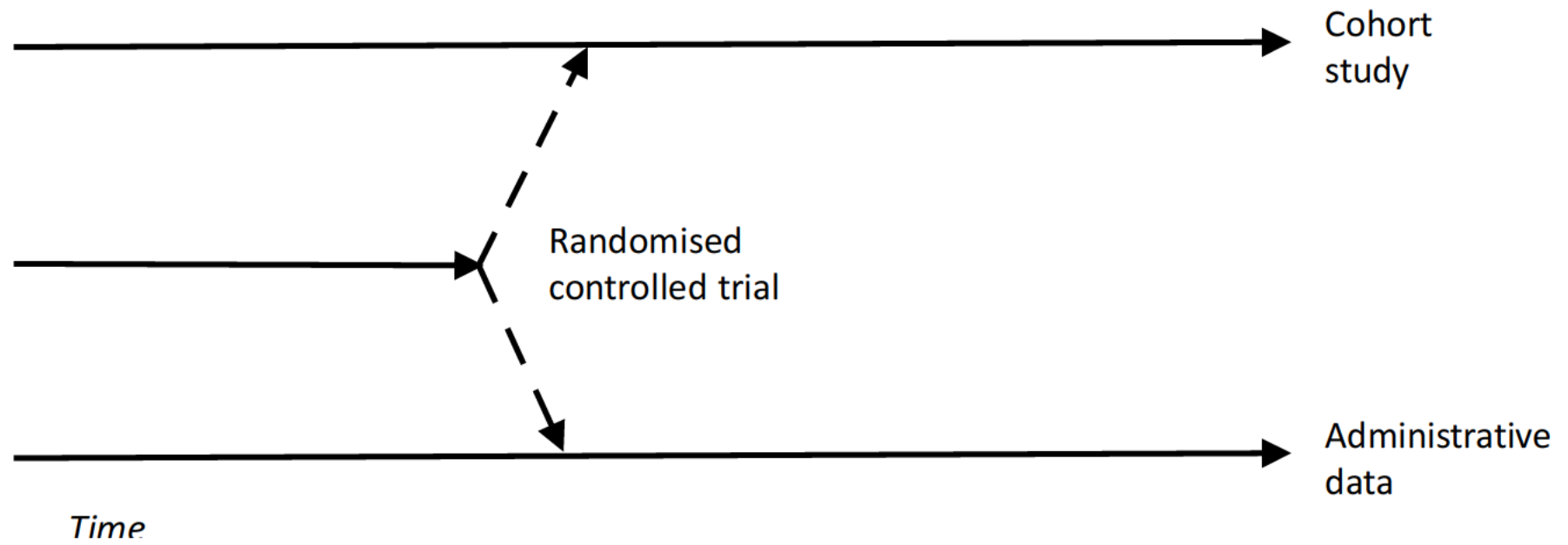


Figure 2.