Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence

C M Booth*1 and I F Tannock2

1Division of Cancer Care and Epidemiology, Queen’s University Cancer Research Institute, 10 Stuart Street, Kingston, ON K7L 5PG, Canada and 2Division of Medical Oncology and Hematology, Princess Margaret Cancer Centre, Toronto, ON, Canada

Recent reports have highlighted the distinct roles and occasional tension between randomised controlled trials (RCTs) and population-based observational research (Concato, 2012; Hershman and Wright, 2012; Goodwin et al, 2013). In this commentary, we will discuss the relative strengths and limitations of each form of research (see Table 1), and propose that well-designed RCTs and population-based observational studies can serve as complementary forms of research to ensure that the results of clinical trials translate into tangible benefits in the general population.

STRENGTHS AND LIMITATIONS OF RCTS

The strength of the RCT rests on its excellent internal validity, which is based largely on the power of randomisation to ensure that the only difference between two treatment arms is their exposure to the treatment of interest. Although randomisation minimises the risk of bias by confounding, there are other biases inherent to RCTs that limit their applicability to the care of patients in routine practice. In particular, patients, providers, and concurrent care in the general population are different from those in clinical trials, and the generalisability (or external validity) of RCTs may be limited (Dans et al, 1998; Meyer, 2010). Although population-based observational research does not enjoy the same level of internal validity at RCTs, well-designed observational studies can offer superior external validity and provide a unique opportunity to evaluate the uptake of new treatments and their outcomes in routine practice.

Most of the substantial improvements in treatment and outcome of patients with cancer over the past four decades have been identified in RCTs. However, patients are highly selected to participate in RCTs, and the greatest limiting factor in interpreting them is that patients seen in routine practice are very different from patients included in RCTs. This is not surprising given that <10% of patients with cancer are entered onto a clinical trial. Patients with advanced age and greater comorbidity (Hutchins et al, 1999; Lewis et al, 2003), and those from lower socioeconomic background (Unger et al, 2013) are under-represented in RCTs. There can also be important differences in the provision of care for patients on RCTs (i.e., highly regulated trial protocols at specialised centres of excellence) compared with patients in routine practice. Given the greater toxicity that is expected when a treatment is applied to a non-selected population with greater comorbidity than subjects included in trials, a small increase in overall (or progression-free) survival observed in a large RCT is likely to disappear when some treatments are applied in routine practice. There is a major difference in generalisability between RCTs that report substantial gain and limited toxicity (e.g., abiraterone for advanced prostate cancer (de Bono et al, 2011), adjuvant trastuzumab for breast cancer (Romond et al, 2005)) and RCTs evaluating drugs with marginal effects and substantial toxicity (e.g., bevacizumab for advanced breast cancer (Miles et al, 2010; Robert et al, 2011), and aflibercept for advanced colorectal cancer (Van Cutsem et al, 2012)). RCTs reporting results of marginal clinical significance to the selected patients recruited to them can be very misleading.

Most oncologists and patients would define a treatment as having benefit if it allows patients to live longer, live better, or both. Unfortunately, a minority of new treatments evaluated in RCTs have achieved these goals. The marked increase in sample size of RCTs provides statistical power to detect treatment differences between arms that are statistically significant but of marginal clinical relevance. RCTs evaluating treatments for cancer are reporting smaller incremental benefits than previously (Booth et al, 2008; Seruga et al, 2010). There is increasing use of surrogate endpoints that have not been validated as predictive of improvement in duration or quality of survival (Booth et al, 2008; Kay et al, 2012) and growing recognition that RCTs...
underestimate and under-report harms from new cancer therapies (Seruga et al, 2011; Niraula et al, 2012).

We and others have described suboptimal reporting of trial findings and various forms of bias associated with disseminating RCT results to practitioners that can adversely influence patient care. Failure to publish studies with negative results can influence results of meta-analyses and treatment guidelines through an imbalanced perspective of the benefits (or lack thereof) of new medical therapies (Krzyszanowska et al, 2003; Tam et al, 2011); hopefully, mandatory trial registration will reduce this bias. Presentation of non-final analyses of RCTs at oncology meetings is commonplace, and in up to 10% of cases the study conclusions will alter substantially between conference presentation and publication of final results (Booth et al, 2009a). Industry-funded RCTs are more likely to be reported as positive than those not sponsored by industry (Djulbegovic et al, 2000; Peppercorn et al, 2007; Booth et al, 2008). As the vast majority of RCTs are now funded by industry (Booth et al, 2008; Kay et al, 2012), clinicians need to recognise the possibility of sponsorship bias. Selective reporting and ‘spin’ can sometimes give the impression that experimental therapy is providing benefit to patients when there is no difference in the primary outcome measure as compared with the control arm and/or increased toxicity associated with the new treatment (Ohorodnyk et al, 2009; Boutron et al, 2010; Altwairgi et al, 2012; Vera-Badillo et al, 2013). We join others in appealing to editors of leading cancer journals to provide a checklist to authors and reviewers to prevent such biased reporting, because it remains prevalent (Saltz, 2008; Booth et al, 2009b).

### STRENGTHS AND LIMITATIONS OF POPULATION-BASED OBSERVATIONAL RESEARCH

Population-based observational studies differ from traditional institutional retrospective studies in that the former include all patients within a given jurisdiction and are therefore less prone to selection and referral biases that plague more traditional forms of observational research.

Large-scale studies have been enabled by advances in computer technology that allow interactions between databases, such as cancer registries or Surveillance, Epidemiology and End Results data, and hospital records. Population-based observational studies can provide information about rare diseases for which there are no RCTs (e.g., small cell cancer of the bladder and adrenocortical cancer; Kerkhofs et al, 2013; Schreiber et al, 2013). Furthermore, changes in the biology (e.g., HPV-related cancer of the oropharynx) and epidemiology of cancer (e.g., gastroesophageal cancer) can be best described using observational research (Devesa et al, 1998; Chaturvedi et al, 2011). Observational studies also provide insights into the care and outcomes of patients under-represented in RCTs, including the elderly and those with comorbidity (Tyldesley et al, 2000; Fossa et al, 2006). Potential risk factors related to developing cancer and the prognostic significance of disease-related characteristics can also be described using observational data.

Given the differences between patients recruited to trials and those seen in routine practice, increased toxicity might be expected when the results of RCTs are applied to routine practice. Population-based studies can provide information about toxicity associated with treatment. Examples include the finding of increased cardiovascular disease and diabetes among men treated with androgen deprivation for localised prostate cancer (Keating et al, 2006), the risks of cardiac disease after radiotherapy for breast cancer (Darby et al, 2013), and long-term toxicities associated with treatment of testicular cancer (Fossa et al, 2007).

Despite compelling evidence from RCTs and published treatment guidelines, physicians may not adopt new medical therapies. Population-based studies can identify gaps in care following

---

**Table 1. An overview of relative strengths and limitations of randomized controlled trials and population-based observational studies**

|                             | Randomised controlled trials                                                                 | Population-based observational studies                                                                 |
|-----------------------------|---------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| **Strengths**               | Excellent internal validity                                                                 | Good external validity                                                                                  |
|                             | Provide precise measures of efficacy and acute toxicity of new therapies under ideal conditions | Provide insight into delivery of care in routine practice to all patients, including elderly and those with comorbidity |
|                             | Because of randomisation, measurement of effect size is less prone to bias                   | Provide information to guide future knowledge translation                                               |
|                             | Allow exploratory measures of secondary endpoints, including patient-reported outcomes and aspects of correlative biology | Can provide evidence of effectiveness of new therapies in the general population                         |
|                             | Can evaluate prognostic and predictive properties of new biomarkers and cancer therapies    | Large samples provide the opportunity to study rare diseases for which RCTs are not possible              |
|                             | Provide a mechanism whereby new (and potentially toxic) treatments can be carefully studied in centres of excellence | Can provide insight into short- and long-term toxicity in routine practice                               |
| **Limitations**             | Limited external validity                                                                   | Limited internal validity: may be difficult to separate effects of a new treatment from other factors    |
|                             | Provide evidence of efficacy (drug effect under ideal circumstances), but not about effectiveness (i.e., true benefit to patients in routine practice) | Population-level databases often do not include detail regarding comorbidity, performance status, and specific treatment plan |
|                             | Applicability to clinical practice can be limited:                                          | Identification of comparative benefit in these studies is prone to multiple biases, including confounding by indication for a given treatment and/or concurrent changes in practice and/or disease biology |
|                             | (i) because patients and practitioners in RCTs are different from those in routine practice |                                                                                                         |
|                             | (ii) elderly and patients with comorbidity are under-represented in RCTs                    |                                                                                                         |
|                             | (iii) often powered to detect a clinically modest effect size that may not apply to less selected patients |                                                                                                         |
|                             | (iv) may use a surrogate primary endpoint that is not a valid measure of patient benefit    |                                                                                                         |
|                             | (v) have limited ability to detect rare and chronic toxicities, especially those that occur in patients with comorbidity or emerge after completion of the trial |                                                                                                         |
The way forward: complementary RCTs and population-based observational studies

We propose that the ideal evolution of evidence for benefit of new cancer therapies would first involve the demonstration of efficacy in a well-designed RCT powered to detect a clinically meaningful benefit. Such trials will need a few hundred patients, but, at least for treatment of incurable cancer, should be smaller than many current trials and have early stopping rules to ensure discontinuation if they have potential to detect a benefit at most minimal differences. Subsequently, population-based observational studies should evaluate patterns of care, toxicity, and the effectiveness of treatment in routine practice. These studies will also provide information needed to improve translation of research findings and quality of care. The use of population-level data is consistent with the CancerLinQ initiative of the American Society of Clinical Oncology to make effective use of ‘big data’.

Patients, clinicians, investigators, and policy makers must surely agree that we need to do more than continue the current trend of conducting yet another mega RCT, demonstrating small (but statistically significant) benefit in highly selected trial patients. If the medical community could optimise the use of current treatments by giving the right treatment to the right patient at the right time, health outcomes in the general population would almost certainly improve far more than the cumulative treatment advances of countless new overpowered RCTs (McGlynn et al., 2003; Woolf and Johnson, 2005). Health services research can be a powerful tool to identify gaps in care and areas for improvement such that we may move towards ‘achieving the achievable’ (Mackillop, 2007). Population-based observational studies and RCTs provide complementary information to improve the lives of patients with cancer, and to provide evidence for and against improvement in outcome at the level of the general population.

Acknowledgements

Dr Booth is supported as a Cancer Care Ontario Research Chair.

Conflict of interest

The authors declare no conflict of interest.

References

Altwairgi AK, Booth CM, Hopman WM, Baetz TD (2012) Discordance between conclusions stated in the abstract and conclusions in the article: analysis of published randomized controlled trials of systemic therapy in lung cancer. J Clin Oncol 30: 3552–3557.

Bach PB, Cramer LD, Warren JL, Begg CB (1999) Racial differences in the treatment of early-stage lung cancer. N Engl J Med 341: 1198–1205.

Baxter NN, Virnig BA, Durham SB, Tuttle TM (2004) Trends in the treatment of early-stage lung cancer. N Engl J Med 341: 443–448.

Birkmeyer JD, Sievers AE, Finlayson EV, Stukel TA, Lucas FL, Batista I, Welch HG, Wennberg DE (2002) Hospital volume and surgical mortality in the United States. N Engl J Med 346: 1128–1137.

The most important limitation is in differentiating between outcomes that are due to adoption of a new treatment and those due to other unrecognised changes in the population under study. Factors that may not be identified or measurable using observational data include stage migration, changes in disease biology, changes in other aspects of management, and confounding by indication. Although statistical modelling techniques such as time series, propensity score, and instrumental variable analyses can mitigate these potential sources of bias, they remain inherent limitations of the study design. However, these limitations do not render this form of research less valuable than insights provided by RCTs, which have their own limitations.

The role of observational studies in defining benefit from treatment can vary depending on the context and quality of the evidence in support of a specific cancer treatment. Where an RCT does not find efficacy of a new cancer treatment, it would be difficult to accept a finding of improved outcomes in an observational study of the general population. In contrast, in situations where RCTs clearly demonstrate efficacy of a new cancer treatment, follow-up observational studies are essential to identify whether practice has changed appropriately, to document harms of therapy in a wider population and in patients of different age and with different comorbidity, and to determine whether patients in routine practice are achieving the expected outcomes. If the observational study finds an improvement in outcome commensurate with results from the RCT, it will support using the treatment in question and the effectiveness of that therapy. Examples of observational studies done in follow-up of landmark RCTs that demonstrated improved survival at the population-level include that evaluating chemoradiotherapy for cervical cancer (Pearcey et al., 2007), the Ontario study describing use of adjuvant chemotherapy for non-small cell lung cancer (Booth et al., 2010), and a study demonstrating the effectiveness of FOLFIRI adjuvant chemotherapy for colon cancer (Sanoff et al., 2012). If outcomes achieved in routine care differ from those expected from an RCT, patient selection and/or delivery of treatment need to be considered carefully. An example was a population-based observational study performed after a large RCT (Pitt et al., 1999) reported benefit of spironolactone for patients with heart failure. The observational data demonstrated that uptake of spironolactone in the real world was associated with increased morbidity and mortality due to hyperkalaemia and no improvement in rates of readmission to hospital for heart failure or overall survival (Juurlink et al., 2004).

Many important clinical questions have not, cannot, and will not be ever addressed in the context of an RCT. In these situations, clinicians rely on information provided by observational research. Oncology practice and policy have been influenced by population-based studies showing that patient outcome is influenced by the interval between surgery and adjuvant chemotherapy for colorectal and breast cancer (Hershman et al., 2006; Lohrisch et al., 2006), hospital and surgeon volume of cancer surgery (Birkmeyer et al., 2002; Derogar et al., 2013), and the extent of lymph node harvest in colorectal cancer (Chen and Bilchik, 2006; Johnson et al., 2006).

Observational studies do have important limitations that must be carefully considered when evaluating treatment benefit. The most important limitation is in differentiating between outcomes that are due to adoption of a new treatment and those due to other unrecognised changes in the population under study. Factors that may not be identified or measurable using observational data include stage migration, changes in disease biology, changes in other aspects of management, and confounding by indication. Although statistical modelling techniques such as time series, propensity score, and instrumental variable analyses can mitigate these potential sources of bias, they remain inherent limitations of the study design. However, these limitations do not render this form of research less valuable than insights provided by RCTs, which have their own limitations.

The role of observational studies in defining benefit from treatment can vary depending on the context and quality of the evidence in support of a specific cancer treatment. Where an RCT does not find efficacy of a new cancer treatment, it would be difficult to accept a finding of improved outcomes in an observational study of the general population. In contrast, in situations where RCTs clearly demonstrate efficacy of a new cancer treatment, follow-up observational studies are essential to identify whether practice has changed appropriately, to document harms of therapy in a wider population and in patients of different age and with different comorbidity, and to determine whether patients in routine practice are achieving the expected outcomes. If the observational study finds an improvement in outcome commensurate with results from the RCT, it will support using the treatment in question and the effectiveness of that therapy. Examples of observational studies done in follow-up of landmark RCTs that demonstrated improved survival at the population-level include that evaluating chemoradiotherapy for cervical cancer (Pearcey et al., 2007), the Ontario study describing use of adjuvant chemotherapy for non-small cell lung cancer (Booth et al., 2010), and a study demonstrating the effectiveness of FOLFIRI adjuvant chemotherapy for colon cancer (Sanoff et al., 2012). If outcomes achieved in routine care differ from those expected from an RCT, patient selection and/or delivery of treatment need to be considered carefully. An example was a population-based observational study performed after a large RCT (Pitt et al., 1999) reported benefit of spironolactone for patients with heart failure. The observational data demonstrated that uptake of spironolactone in the real world was associated with increased morbidity and mortality due to hyperkalaemia and no improvement in rates of readmission to hospital for heart failure or overall survival (Juurlink et al., 2004).

The most important limitation is in differentiating between outcomes that are due to adoption of a new treatment and those due to other unrecognised changes in the population under study. Factors that may not be identified or measurable using observational data include stage migration, changes in disease biology, changes in other aspects of management, and confounding by indication. Although statistical modelling techniques such as time series, propensity score, and instrumental variable analyses can mitigate these potential sources of bias, they remain inherent limitations of the study design. However, these limitations do not render this form of research less valuable than insights provided by RCTs, which have their own limitations.

The role of observational studies in defining benefit from treatment can vary depending on the context and quality of the evidence in support of a specific cancer treatment. Where an RCT does not find efficacy of a new cancer treatment, it would be difficult to accept a finding of improved outcomes in an observational study of the general population. In contrast, in situations where RCTs clearly demonstrate efficacy of a new cancer treatment, follow-up observational studies are essential to identify whether practice has changed appropriately, to document harms of therapy in a wider population and in patients of different age and with different comorbidity, and to determine whether patients in routine practice are achieving the expected outcomes. If the observational study finds an improvement in outcome commensurate with results from the RCT, it will support using the treatment in question and the effectiveness of that therapy. Examples of observational studies done in follow-up of landmark RCTs that demonstrated improved survival at the population-level include that evaluating chemoradiotherapy for cervical cancer (Pearcey et al., 2007), the Ontario study describing use of adjuvant chemotherapy for non-small cell lung cancer (Booth et al., 2010), and a study demonstrating the effectiveness of FOLFIRI adjuvant chemotherapy for colon cancer (Sanoff et al., 2012). If outcomes achieved in routine care differ from those expected from an RCT, patient selection and/or delivery of treatment need to be considered carefully. An example was a population-based observational study performed after a large RCT (Pitt et al., 1999) reported benefit of spironolactone for patients with heart failure. The observational data demonstrated that uptake of spironolactone in the real world was associated with increased morbidity and mortality due to hyperkalaemia and no improvement in rates of readmission to hospital for heart failure or overall survival (Juurlink et al., 2004).
