STATISTICAL AND METHODOLOGICAL ISSUES IN EVALUATION OF

INTEGRATED CARE PROGRAMS

By

CHENGLIN YE, B.Sc. (Honours), M.Sc. (Statistics)

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

McMaster University DOCTOR OF PHILOSOPHY (2014) Hamilton, Ontario

(Health Research Methodology – Biostatistics Specialization)

TITLE: Statistical and Methodological Issues in Evaluation of Integrated Care Programs

AUTHOR: Chenglin Ye B.Sc. (McMaster University), M.Sc. (McMaster University)

SUPERVISOR: Professor Lehana Thabane

NUMBER OF PAGES: ix, 135

ABSTRACT

Background

Integrated care programs are collaborations to improve health services delivery for patients with multiple conditions.

Objectives

This thesis investigated three issues in evaluation of integrated care programs: (1) quantifying integration for integrated care programs, (2) analyzing integrated care programs with substantial non-compliance, and (3) assessing bias when evaluating integrated care programs under different non-compliant scenarios.

Methods

Project 1: We developed a method to quantity integration through service providers' perception and expectation. For each provider, four integration scores were calculated. The properties of the scores were assessed.

Project 2: A randomized controlled trial (RCT) compared the Children's Treatment Network (CTN) with usual care on managing the children with complex conditions. To handle non-compliance, we employed the intention-to-treat (ITT), as-treated (AT), per-protocol (PP), and instrumental variable (IV) analyses. We also investigated propensity score (PS) methods to control for potential confounding.

Project 3: Based on the CTN study, we simulated trials of different non-compliant scenarios. We then compared the ITT, AT, PP, IV, and complier average casual effect methods in analyzing the data. The results were compared by the bias of the estimate, mean square error, and 95% coverage.

Results and conclusions

Project 1: We demonstrated the proposed method in measuring integration and some of its properties. By bootstrapping analyses, we showed that the global integration score was robust. Our method has extended existing measures of integration and possesses a good extent of validity.

Project 2: The CTN intervention was not significantly different from usual care on improving patients' outcomes. The study highlighted some methodological challenges in evaluating integrated care programs in a RCT setting.

Project 3: When an intervention had a moderate or large effect, the ITT analysis was considerably biased under non-compliance and alternative analyses could provide unbiased results. To minimize the bias, we make some recommendations for the choice of analyses under different scenarios.

PREFACE

This thesis is a "sandwich thesis", which combines three individual projects prepared for publication in peer-reviewed journals. The following are the contributions of Chenglin Ye in all the papers included in this dissertation: conceiving the research ideas, developing research questions, designing the studies, developing the analysis and simulation plans, conducting all statistical analysis and simulations, producing all figures and tables, writing the manuscripts, submitting the manuscripts, and responding to reviewers' comments. My co-authors contributed to the acquisition of the datasets, provision of clinical expertise, and critical revision of the manuscripts. The work in this thesis was conducted between September 2010 and January 2014. The first and the second papers have been published, and the third paper has been submitted to a peer-reviewed journal.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis supervisor, Dr. Lehana Thabane. He has consistently been a supportive mentor, inspiring supervisor, and encouraging friend throughout my graduate studies at McMaster University. I am also grateful to Drs. Gina Browne and Joseph Beyene for their invaluable and timely contributions to the research presented in this thesis.

I would like to thank Drs. Deborah Cook, Gary Foster, Lisa Dolvich, and Eleanor Pullenayegum for collaborations on many exciting projects. Many thanks to all my friends and colleagues from McMaster University community and St. Joseph's Healthcare Hamilton for their help and encouragement.

I am forever indebted to my beloved mon, Guimei Yang for her unconditional love and support.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In 2010, the total Canadian health care expenditures were approximately $191.6 billons [1]. It is forecasted that six Canadian provinces will spend more than half of their total revenues on health care by the year 2020 if the current trend continues [2]. The aging of populations and growing prevalence of chronic conditions are two main factors contributing to the escalating costs [3-5]. Patients with chronic illness or functional disabilities consume most health services [6,7] since they often endure complex conditions and each condition requires specific and continuous care. Traditional way of delivering health services for those populations has been fragmented. For instance, a patient with multiple conditions is referred to different specialists. Different practitioners are responsible for the same patient but the services provided are disjointed without a single plan to address what the patient needs overall. This fragmented approach creates duplicative diagnosis and treatment for the patient and potentially delay the optimal timing for appropriate care. Also, providing health services in a non-active and discontinuous fashion will be inefficient in managing patients with complex needs and further increase the burden of health care system. Under limited resources, developing a way of delivering proactive, comprehensive, continuous, and patient-centered care to address target patients' multiple needs has emerged as a central topic.

Over the last decade, integration has been advocated as a viable strategy for improving the organization of health services. A prominent hypothesis is that integration of health services will lead to higher quality of care at a lower cost and maintain or improve patients' health and satisfaction [8,9]. Integration is defined in a variety of contexts and the majority of integration literature focuses on primary care, hospital care, or mental health care [10,11]. In general, integrated care is "patients care that coordinated across professionals, facilities, and support systems; continuous over time and between visits; tailored to the patients' needs and preferences; and based on shared responsibility between patient and caregivers for optimizing health [12]."

Integrated care programs have been primarily evaluated by qualitative tools. Only a few methods for quantifying the degree of integration exist. A systematic review [8] identified 18 quantitative measures of integration and each one had a different methodological approach. However, none of those measures provided a single score of integration, making it difficult to relate the degree of integration to patients' outcomes. Despite promising results, some randomized controlled trials (RCTs) designed to evaluate integrated care programs failed to show significant effects on patients' outcomes [13-15]. Those trials were largely limited by non-compliant rates and settings of the studies. Under those limitations, the conclusion of a study might be sensitive to the choice of analysis and the extent of bias incurred in estimating treatment effects was unclear.

The objective of this thesis is to address some issues around evaluating integrated care programs, through the lens of measurement development, sensitivity analyses and simulations, and to provide some directions for future research. Three specific issues are investigated: (1) the quantification of the degree of integration for an integrated service network, (2) the sensitivity analysis of an integrated care program by different statistical methods, and (3) the assessment of bias in estimating effects of integrated care programs under different non-compliant scenarios.

Issue 1: Measuring integration of an integrated service network

Studies have shown that integrating local services improves patient outcomes, for example, reducing functional decline in the elderly [16], preventing avoidable hospitalization [17], and minimizing complications for diabetes patients [18]. Most researchers support the view that coordinating available, necessary, and preferable services to patients help achieve target outcomes more easily with less total costs [19].

However, the extent to which health services are integrated in a health service network is often unknown and rarely measured. Without a proper measurement, integration efforts can be hardly confirmed and monitored. There are two key limitations in the existing measures [8]. Firstly, the existing measures generally do not distinguish service providers' perception of integration from their expectation. The goal of integration is to coordinate health services to meet patients' multiple needs by aligning different services instead of merely pooling resources. Both perception and expectation among all service

providers are essential in such alignment. Thus, by matching perception with expectation, a measure of integration ideally will show how well a service provider is engaging at the level that all other partners expected him to be. Secondly, the existing measures commonly assume that a higher level of involvement means a better integration. This can be misleading because a high level of input by a service provider is not necessarily an agreed-upon level of involvement by other providers towards a common goal. A measure of integration needs to reflect an agreement of involvement among all service providers. Developing a measure that addresses those issues will contribute importantly to the quantitative methods of measuring integration.

Issue 2: Analyzing the Children's Treatment Network trial

Children with special health care needs (CSHCN) are those "who have or are at increased risk for a chronic physical, developmental, behavioral, or emotional condition, and who also require health and related services of a type of amount beyond that required by children generally" [20]. In Canada, 13% - 18% of children are considered to have special needs [21]. Two thirds of CSHCN are unable to get necessary treatments and more than one third of their parents do not know where to look for help [22]. Since some CSHCN requires specialized services that are not available locally, their parents have to repeatedly travel outside the region to get those services for them.

In responding to the problems, the Children's Treatment Network (CTN) model of care was launched in 2006, which targeted the children with complex health needs in Simcoe

County and York Region, Ontario Canada. The CTN coordinated local service providers to provide a single point of service access and adopted a multidisciplinary team approach to deliver comprehensive care to address target children's multiple needs.

A RCT was conducted to compare the CTN integrated care with usual care from 2007 to 2009 [23]. The primary outcome was the improvement of children's psychosocial function at 2 years. The result did not show a significant effect of the CTN integrated care and this finding was largely limited by a substantial non-compliant rate. About 52% of the children in the intervention group only received parts of the CTN integrated care. Suboptimal intervention fidelity at the early development stage partly contributed to the non-compliance. However, it is unclear how robust the analysis was and how the result would change if alternative methods of dealing with non-compliance were used. A sensitivity analysis by employing different methods of handling non-compliance will provide a better insight into the real effect of CTN integrated care.

Issue 3: Assessing the bias of estimates under different non-compliant scenarios

Patients' non-compliance is a common problem in RCTs. Non-compliance undermines randomization and the 'fair' comparison between treatment groups, potentially leading to a biased estimate of treatment effect. While different analyses have been proposed to deal with non-compliance, the extent of bias in estimated treatment effect is rarely studied. The interpretation of an analysis also varies depending on the nature of non-compliance and the objective of a study. In the context of RCTs evaluating integrated care programs,

non-compliance may be a practical issue since those trials are generally conducted in real-life settings and participants are much less restricted.

Intention-to-treat (ITT), as-treated (AT), and per-protocol (PP) analyses are commonly used in analyzing non-compliant data. An ITT analysis is considered as the standard for RCTs but it has been recommended to use different methods to analyze RCTs with substantial non-compliance [24]. A literature review [25] randomly selected 100 RCTs published in high impact journals in 2008. Out of 98 RCTs which reported non-compliance, 46 employed variations of PP analyses in addition to an ITT analysis, 5 employed other types of analyses, and 47 did not implement some statistical method to address non-compliance. Another class of methods to deal with non-compliance includes instrumental variable (IV) and complier average causal effect (CACE) analyses [26-29]. Very few studies have compared those methods of analyzing non-compliant data by the bias of the estimates, mean square error, and 95% coverage of the true value. Bang and Davis [30] compared ITT, AT, PP, and IV methods. They showed that ITT and IV analyses were biased in certain non-compliant cases but did not include CACE analysis in the comparison. Also, the situation where there is no crossover between treatment groups was not considered. That scenario is common when a new intervention is only accessible to patients who were offered it. McNamee [31] also compared ITT, AT, PP, and IV analyses and concluded that an ITT analysis was not always biased towards the null while AT and PP analyses were generally biased. Sheng and Kim [32] investigated the effect of non-compliance on ITT analysis of equivalence trials and showed that non-compliance

did not always favor the null hypothesis, i.e. no difference between treatment groups. Hertogh et al. [33] concluded that the IV method could give insight into confounding by non-compliance in RCTs. In addition, different factors may characterize patients' non-compliant behaviours. For example, non-compliers may always receive or reject a new intervention regardless of their treatment assignment. Patients with certain characteristics may also tend to comply more than the others. For some cases, patients may receive parts of the intervention even they have not fully complied with it. A further investigation on common analyses of non-compliance will contribute to the understanding on the bias of estimated treatment effects and optimal choice of analyses under different non-compliant scenarios.

Scope of the Chapters

This is a sandwich thesis of three papers. Each paper matches with one of the issues described above. The papers are separated into three chapters beginning with Chapter 2.

Chapter 2 describes the issue of measuring integration. We developed a method to calculate integration scores for each service provider of an integrated service network and a global integration score of the whole network. The method was built on the Human Service Integration Measure Scale developed by Browne et al. [10]. Each service provider rated the perceived and expected level of partnership with every other service provider on a 5-point ordinal scale. The ratings were sorted into four types: self-perceived, self-expected, group-perceived, and group-expected ratings. Two matrices were developed to

facilitate the calculations and the details are explained in Chapter 2. We quantified the integration by calculating the agreement between: group-perceived and group-expected ratings; self-perceived and group-expected ratings; group-perceived and self-expected ratings; and self-perceived and self-expected ratings. The integration scores were reported in percentage with 95% confidence interval. We also used a graphical display to summarize the overall integration and calculated a global score. The sensitivity of the global integration score was examined by bootstrapping methods. The proposed method was then applied in measuring the degree of integration among service providers of the CTN.

Chapter 3 describes with the sensitivity analysis of the CTN trial. We compared the ITT with alternative approaches of analyzing the CTN data, including the AT, PP, and IV methods. The AT analysis compared the children who received complete CTN integrated care with the ones who received only parts or none of it. The PP analysis compared the children who fully complied with their assigned treatment. We employed propensity score (PS) methods to adjust for potential prognostic imbalance in the AT and PP analyses since both analyses compromised original randomization. Four PS methods were used: matching on PS, stratifying on PS, weighting on PS, and adjusting for PS as a covariate. The IV analysis used the randomization allocation as the instrumental variable to estimate the causal effect of CTN integrated care on target children's outcome. We applied the Delta method to calculate the standard error of the IV estimate. All results were reported in estimated effect size (expressed as the mean difference between groups), 95%

confidence interval, and p value. The estimates were compared by their magnitude, precision, direction, and significance.

Chapter 4 describes the assessment of bias by applying different methods to analyze non-compliant scenarios. We simulated hypothetical RCTs by using the key parameters estimated from the CTN trial to mimic a similar study setting. Different non-compliant scenarios were generated by varying three factors: the type of non-compliers, the randomness of non-compliance, and the degree of non-compliance. In total, we generated 30 different non-compliant scenarios. We analyzed each of those scenarios by the ITT, AT, PP, IV, and CACE analyses and compared the results by the bias of the estimate, mean square error, and 95% coverage of the true value. A sensitivity analysis was also conducted to examine the impact of dichotomizing patients' adherence to treatment. Two cut-offs were investigated: 80% (i.e. a patient considered to have fully complied if he or she had received 80% of all elements of the treatment) and 100%.

Chapter 5 summarizes the key findings of Chapters 2 to 4, and discusses the implications and limitations of the thesis. The common goal of all three papers is to advance our understanding on the statistical and methodological issues around the evaluation of integrated care programs. Results of the individual projects will provide knowledge for designing future studies to evaluate similar interventions.

References

1. Canadian Institute for Health Information. Health care in Canada 2010 Canada: CIHI; 2010 Dec. Available from: https://secure.cihi.ca/estore/productSeries.htm?pc=PCC64. Accessed 21 May 2012.

2. Skinner BJ, Rovere M. Paying more, getting less: measuring the sustainability of government health spending in Canada 2009 Report. Canada: Fraser Institute; 2009 Nov. Available from: http://www.ihe.ca/publications/health-db/geo/448/. Accessed 21 May 2012.

3. The Conference Board of Canada. Understanding health care cost drivers and escalators Ottawa; 2004 Mar. Available from: http://www.health.alberta.ca/documents/Health-Costs-Drivers-CBC-2004.pdf. Accessed 21 May 2012.

4. Christensen K, Doblhammer G, Rau R, et al. Ageing populations: the challenges ahead. The Lancet 2009;374(9696):1196–208.

5. Canadian Health Services Research Foundation. Myth: Canada's system of healthcare financing is unsustainable Ottawa: CHSRF; 2007 Jan. Available from: http://www.chsrf.ca/PublicationsAndResources/Mythbusters/ArticleView/07-12-01/bb0d6092-9ec8-45c0-a38d-de696b7b25c9.aspx. Accessed 21 May 2012.

6. Denton FT, Spencer BG. Chronic health conditions: changing prevalence in an aging population and some implications for the delivery of health care services. Canadian Journal on Aging 2010;29(1):11–21.

7. Inkelas M, Raghavan R, Larson K, et al. Unmet mental health need and access to

services for children with special health care needs and their families. Ambulatory Pediatrics 2007;7(6):431–8.

8. Strandberg-Larsen M, Krasnik A. Measurement of integrated healthcare delivery: a systematic review of methods and future research directions. International Journal of Integrated Care [serial online] 2009 Feb 4; 9. Available from: http://www.ijic.org/index.php/ijic/article/view/305/608. URN:NBN:NL:UI:10-1-100547. Accessed 21 May 2012.

9. Fleury M, Mercier C. Integrated local networks as a model for organizing mental health services. Administration and Policy in Mental Health 2002;30(1):55–73.

10. Browne G, Roberts J, Gafni A, et al. Conceptualizing and validating the human services integration measure. International Journal of Integrated Care [serial online] 2004 May 19; 4. Available from: http://www.ijic.org/index.php/ijic/article/view/98/195. URN:NBN:NL:UI:10-1-100340. Accessed 21 May 2012.

11. Topping S, Calloway M. Does resource scarcity create interorganizational coordination and formal service linkages? A case study of a rural mental health system. Advances in Health Care Management 2000;1(1):393–419.

12. Singer SJ, Burgers J, Friedberg M. Defining and measuring integrated patient care: promoting the next frontier in health care delivery. Medical Care Research and Review 2011;68(1):112-127.

13. Ye C, Browne G, Beyene J et al. A sensitivity analysis of the Children's Treatment Network trial: a randomized controlled trial of integrated services versus usual care

for children with special health care needs. Clinical Epidemiology 2013;5(1):373-385.

14. Bogner HR, Morales KH, de Vries HF, et al. Integrated management of type 2 diabetes mellitus and depression treatment to improve medication adherence: a randomized controlled trial. Annals of Family Medicine 2012;10(1):15-22.

15. Markle-Reid M, Browne G, Gafni A, et al. The effects and costs of a multifactorial and interdisciplinary team approach to falls prevention for older home care clients 'at risk' for falling: a randomized controlled trial. Can J Aging 2010;29(1):139-161.

16. Bernabel R, Landl F, Ganbassi G, et al. Randomised trial of impact of model of integrated care and case management for older people living in the community. British Medical Journal 1998;316(7141):1348–51.

17. Beland F, Bergman H, Lebel P, Clarfield AM, Tousignant P, Contandriopoulos AP, et al. A system of integrated care for older persons with disabilities in Canada: results from a randomized controlled trial. Journal of Gerontology Series A: Biological Sciences and Medical Sciences 2006;61(4):367–73.

18. Larsen DL, Cannon W, Towner S. Longitudinal assessment of a diabetes care management system in an integrated health network. Journal of Managed Care Pharmacy 2003;9(6):552–8.

19. Browne G, Byrne C, Robert J, Gafni A, Whittaker S. When the bough breaks: provider-initiated comprehensive care is more effective and less expensive for sole support parents on social assistance. Social Science and Medicine 2001;53(12):1697–710.

20. McPherson M, Arango P, Fox H, et al. A new definition of children with special

health care needs. Pediatrics 1998;102(1):137-139.

21. Bethell CD, Read D, Blumberg SJ, et al. What is the prevalence of children with special health care needs? Toward an understanding of variations in findings and methods across three national surveys. Matern Child Health J 2008;12(1):1-14.

22. Cossette L, Doclos E. A profile of disability in Canada 2001 (PALS). Ministry of Industry: Statistics Canada 2001. Available from: http://www.statcan.gc.ca/pub/89-577-x/. Accessed May 20, 2013.

23. Thurston S, Paul L, Browne G, et al. Early results: the 2-year comparative effect and expense of more and less integration of services that provide treatment and rehabilitation for children with multiple disabilities: a randomized controlled trial. Children's Treatment Network of Simcoe York, Final research report, March 2010.

24. Hernán MA, Hernández-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. Clinical Trials 2012, 9(1):48-55.

25. Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. Trials 2012, 13:84.

26. Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. Statistical Methods in Medical Research 2005, 14(4):369-395.

27. Martens EP, Pestman WR, de Boer A, et al. Instrumental variables: application and limitations. Epidemiology 2006, 17(3):260-267.

28. Jo B. Estimation of intervention effects with noncompliance: Alternative model specifications. Journal of Educational and Behavioral Statistics 2002, 27(4):385-409.

29. Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. Biometrics 2009, 65(2):640-649.

30. Bang H, Davis CE. On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? Statistics in medicine 2007, 26(5):954-964.

31. McNamee R. Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. Statistics in Medicine 2009, 28(21):2639-2652.

32. Sheng D, Kim MY. The effects of non-compliance on intent-to-treat analysis of equivalence trials. Statistics in Medicine 2006, 25(7):1183-1199.

33. Hertogh EM, Schuit AJ, Peeters PH, et al. Noncompliance in lifestyle intervention studies: the instrumental variable method provides insight into the bias. Journal of Clinical Epidemiology 2010, 63(8):900-906.

CHAPTER 2

Measuring the degree of integration for an integrated service network

Chenglin Ye[1], Gina Browne[2], Valerie S Grdisa[3], Joseph Beyene[1], Lehana Thabane[1*]

Affiliations:

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton ON, L8S 4L8 Canada

[2]School of Nursing, McMaster University, Hamilton ON, L8S 4L8 Canada

[3]Healthcare advisory, Management consulting, KPMG Canada, Toronto ON, Canada

[*]Corresponding author:

Lehana Thabane

Email: thabanl@mcmaster.ca

Phone: +1/905-522-1155 ext.33720

Fax: +1/905-308-7212

Abstract

Background: Integration involves the coordination of services provided by autonomous agencies and improves the organization and delivery of multiple services for target patients. Current measures generally do not distinguish between agencies' perception and expectation. We propose a method for quantifying the agencies' service integration. Using the data from the Children's Treatment Network (CTN), we aimed to measure the degree of integration for the CTN agencies in York and Simcoe.

Theory and Methods: We quantified the integration by the agreement between perceived and expected levels of involvement and calculated four scores from different perspectives for each agency. We used the average score to measure the global network integration and examined the sensitivity of the global score.

Results: Most agencies' integration scores were less than 65%. As measured by the agreement between every other agency's perception and expectation, the overall integration of CTN in Simcoe and York was 44% (95% CI: 39% - 49%) and 52% (95% CI: 48% - 56%), respectively. The sensitivity analysis showed that the global scores were robust.

Conclusion: Our method extends existing measures of integration and possesses a good extent of validity. We can also apply the method in monitoring improvement and linking integration with other outcomes.

Keywords: integration measure, perception, expectation, collaboration agreement

Background

Spending on healthcare in Canada continues to outpace government revenue and economic growth [1]. In 2010, the total Canadian healthcare expenditures were approximately $191.6 billion [2]. Fraser Institute, a Canadian think-tank, forecasts that if the recent trend continues, six of 10 provincial governments will spend more than half of total revenue on healthcare by the year 2020 [3]. Many factors contribute to escalating healthcare costs that include: development of new drugs, availability of expensive health services, and an increase in co-morbid chronic conditions in our aging populations [4-6]. The sustainability of the existing Canadian universal care system is a growing concern [7].

In general, patients with chronic illness or functional limitations are the major consumers of community health services [8,9]. For example, in addition to taking psychopharmacologic drugs, patients with mental illness would need psychiatric counseling, rehabilitative therapy, and other health and social services to help them and their families. Usually, such primary care services are delivered by community health centres—which are operated autonomously [10]. Patients often receive help for a single problem at a time and endure duplicative processes to receive multiple services. With limited resources, we can hardly meet the total needs of a population by providing health services in an isolated and fragmented fashion. As a result, patients may experience prolonged wait-times and courses of treatment.

Over the last decade, integration has been advocated as a viable strategy for improving the organization and delivery of health services. A majority of the literature on integrating services discusses the integration of health services on primary care, hospital care, or mental health care [8-11]. Often, the integration of health services is framed under the notion of "continuity of care" [12]. A prominent hypothesis is that integration of health services leads to higher quality of care at a lower cost and maintains or improves patients' health and satisfaction [13,14]. In this paper, we adopt the definition introduced by Browne et al. who define integration as the coordination of a comprehensive spectrum of services (e.g., health, education, community and social services) provided by multiple agencies [10]. We use providers and agencies interchangeably to represent the publicly funded organizations that deliver community health services.

The integrated approach of delivering services has potential advantages. Patients will receive comprehensive care and are likely to gain a much better outcome. Another advantage is that service providers can improve their caseload by coordinating with other providers and reducing duplication, and thus, better meet the overall needs of a population. There are different ways to initiating integration within a community. One is by legislative change such as the Health Action Zones in the United Kingdom whereby a whole geographic area receives the total funding for public and health services. The implementation of Local Health Integration Networks (LHINs) in Ontario, Canada is another example where local region receives funds for acute, community, and long-term care services. Regional health authorities develop service agreements with performance

indicators. Savings are kept within the region and used for other priorities. The other approach begins at a local level whereby agencies collaborate together to serve more clients without increasing funding allocation. The local community develops own collaboration in service delivery. The Children's Treatment Network (CTN) in Ontario, Canada is an example of the third approach, for serving the children with complex health problems.

Studies have shown that integrating local services improves patient outcomes, such as reducing functional decline in the elderly [15], preventing avoidable hospitalizations [16], and minimizing the risk of developing diabetes-related complications [17]. One sentence summarizes the current evidence: healthcare providers can achieve target outcomes more easily with less investment by coordinating available, necessary, and preferable human services to patients [18]. From a societal perspective, proactive and comprehensive services are more effective and less expensive because giving people what they need in a coordinated fashion results in a reduced use of other services [19]. The emphasis on human services is the result of accumulating evidence that the factors (besides genetic predisposition) determining health are social, environmental, educational, and personal in nature. However, the extent to which those services are integrated for addressing patients' health needs is often unknown and rarely measured.

Without quantifying the degree of integration, the effect of collaborative effort can be hardly identified. Without measuring and monitoring the actual integration, it is difficult

for network planners to make decisions and confirm successful implementation [20].

Thus, we need a valid method for measuring the degree to which a network has achieved

in integrating services. A recent systematic review [13] identified 18 quantitative

measures of integration. None of those measures separate agencies' current perception of

involvement from their expectation. Since agencies would differ in the level of

involvement necessary to meet patients' needs, a rational measure should reflect the

agreement between perceived and expected levels of involvement. In marketing,

researchers have used the gap between the expected performance and the perceived

experience as an objective measure of customer satisfaction [21]. Current measures

generally do not distinguish between agencies' perception and expectation on integration

although separate measures offer an appropriate and reliable way to identify the 'gap' in

integration [22]. Browne et al. develop the Human Service Integration Measure (HSIM)

scale [10] that includes separate measures of observed and expected integrations. Another

key limitation in existing measures is the common assumption that a higher level of

involvement is always a better integration. A high level of input by an agency is not

necessarily an agreed-upon involvement by all agencies [10,13,23]. A measure of

integration needs to account for the expectations of all underlying agencies in their

pursuit of a common goal to improve patient outcomes [24-26]. In this paper, we

introduce a new method to measure the degree of service integration among agencies. Our

proposed method has unique strengths by: quantifying the gap between perceived and

expected levels of involvement within a network; creating a relational score by the

agreement among all agencies; measuring the integration from four different perspectives;

and creating integration scores for individual agencies and global integration scores for

the entire network as a whole. In this paper, we use the words 'perceived', 'observed',

and 'actual' interchangeably to mean the current level of involvement perceived by

agencies. We use the words 'expected' or 'optimal' to mean the level of involvement

expected by agencies. Using the data from the Children's Treatment Network (CTN), our

objectives are to: 1) measure the degree of integration among agencies of CTN; 2)

calculate global integration scores of CTN; and 3) assess the sensitivity of the global

integration score based on different approaches for estimating the global integration

score.

Theory and Methods

Study design and study population

The data were drawn from a cross-sectional study that evaluated the integration of CTN

in 2006, one year after its inception. The study measured the degree of integration among

agencies of CTN in Simcoe and York separately.

In Simcoe County and York Region of Ontario, Canada, families with severely disabled

child(ren) had limited access to specialized treatments and had to travel outside their

region to receive services. They used services in a disjointed and self-directed manner and

often received suboptimal outcomes. In response to these problems, Simcoe County and

York Region launched the CTN in 2005. Targeting the children with complex health

problems, local service agencies in CTN collaborated together to deliver comprehensive

therapeutic and psychosocial services to the children and their families. This innovative

and proactive model integrated existing service agencies – including health, recreational,

educational, social, mental health, and community resources (*www.ctn-simcoeyork.ca*).

Each family served by CTN had a unique team of service providers for a long-term basis.

The interdisciplinary team provided a single point of contact, health assessment, service

coordination, and a comprehensive plan of care for the children.

Characteristics of the CTN agencies

At the time of measurement, there were 27 and 36 agencies in Simcoe and York,

respectively. The CTN represented health, educational, social, justice, recreational, and

cultural sectors. There were different types of agencies in each sector that included: early

years, Healthy Babies, adolescent support, rehabilitation, home care, social assistance,

child protection, mental health, recreation, leisure services, etc.

The Human Service Integration Measure Scale

To quantify the level of partnership involvement, we used the latest version of the Human

Service Integration Measure (HSIM) scale, a 5-point ordinal scale developed by Browne

et al. to validate their integration framework [10]. Representatives of agencies can fill out

the measures by web-form, phone, or in-person interview. An interviewer will ask each

agency to rate its current and expected levels of involvement with the other agencies

within the network. An example of the measure is provided in Figure 1. For agencies that

have more than one representative, the interviewer will average the ratings from all representatives and round it to the nearest integer.

Organizing the data for calculating integration

We have developed two square matrices to organize the responses of perceived involvement and expected involvement, respectively. Each column in the matrix contains an agency's ratings on every other agency (i.e. self-ratings) and each row contains other agencies' ratings on the same agency (i.e. group-ratings). An illustrative example is provided in Figure 2. In the matrix of perceived involvement, the first column contains the level of involvement that Agency A has perceived with every other agency (e.g. 1, 1, and 1) and the first row contains the level of involvement that other agencies have perceived with Agency A (e.g. 2, 3 and 4). Similarly, in the matrix of expected involvement, the first column contains the level of involvement that Agency A has expected with every other agency (e.g. 1, 2, and 3) and the first row contains the level of involvement that other agencies have expected with Agency A (e.g. 1, 1, and 1). The ratings for other agencies are organized in the same way. Every agency has four types of ratings, namely, self-perceived, self-expected, group-perceived, and group-expected ratings.

Calculating the agency integration scores

We measured an agency's integration score by the agreement between perceived and expected involvements with other agencies and defined the agreement as the percentage

of pairs of agreed perceived and expected ratings. Thus, our integration score was an agreed-upon score on the level of involvement among agencies. For example, if there were 80% of agencies whose group-perceived scores on Agency X were same as their group-expected scores, then the corresponding degree of integration for Agency X would be 80%. Our integration framework measured the agreement from four perspectives shown in Figure 3, including the agreement between: the group-perceived and group-expected involvements (P1), the self-perceived and group-expected involvements (P2), the group-perceived and self-expected involvements (P3), and the self-perceived and self-expected involvements (P4).

Calculating the global integration score

We estimated the global integration score of a network by the average integration score. As shown in Figure 4, the graded area represented the global integration score of a network based on scores from all the network agencies. The blank area on the diagram indicates the gap in the degree of global network integration. The total area is the sum of the graded and the blank areas of the diagram which equals 1 (i.e. 100% integration). Essentially, the graded area represents the average integration score. We calculated corresponding global integration score with 95% confidence interval (CI)) as the graded area based on P1, P2, P3, and P4 integration scores.

Assessing the sensitivity of global integration score

We conducted a sensitivity analysis by comparing with other methods for estimating the

global integration score: the weighted-average method and the bootstrap method. In the

weighted-average method, we attached a different weight (w) to each agency based on the

variance of its integration score. This method adjusted for the precision of an estimated

integration score. An agency with a more precise integration score contributed more to

the global score for the whole network. For the bootstrapping method, we used three

different bootstrap procedures: the standard, the balanced, and the Bayesian procedures.

Bootstrap is a common resampling method for improving estimation and confidence

intervals of an unknown parameter [27-29]. Different procedure requires a different

resampling algorithm and thus, estimates the sampling error differently. The standard

bootstrap can produce a bias-corrected estimate [29] that largely reduces the potential bias

arising in estimation. The balanced bootstrap is similar to the standard procedure but

bootstrap samples are balanced. Compared with the standard bootstrap, the balanced

bootstrap generally improves the efficiency of simulation [30]. The Bayesian bootstrap

uses a different algorithm and approximates a posterior distribution of the global score

instead of a sampling distribution [31]. For both the standard and balanced bootstrap

procedures, we computed the 95% bias-corrected and accelerated CI that adjusted both

bias and skewness in bootstrap sampling [28]. For the Bayesian bootstrap procedure, we

computed the 95% credibility interval (CrI) instead. The details of calculations and

procedures were provided in the Appendix. We performed all statistical analyses in the

software package R version 2.12.1.

Results

The integration scores for agencies of CTN

We summarized the integration scores for CTN agencies in Table 1. For confidentiality,
we kept agencies anonymous and labeled them by Arabic numbers. The response rate was
89% in Simcoe and 64% in York. All agencies (i.e. both respondents and non-
respondents) received a P1 score. However, only respondents received P2, P3, and P4
scores because those calculations required the agencies' ratings on other agencies. In
Simcoe, P1, P2, P3, and P4 integration scores varied from 25% to 82%, 4% to 83%, 13%
to 83%, and 4% to 96%, respectively. In York, P1, P2, P3, and P4 integration scores
varied from 27% to 73%, 27% to 81%, 9% to 82%, and 23% to 97%, respectively.

The global integration scores for the CTN

Global integration scores of CTN were summarized in Table 2. In Simcoe, P1, P2, P3,
and P4 global integration scores with 95% CI were 44% (39%, 49%), 43% (36%, 51%),
43% (35%, 52%), and 44% (32%, 55%), respectively. In York, they were 52% (48%,
56%), 54% (48%, 61%), 54% (45%, 63%), and 52% (43%, 61%), respectively. The
global integration of CTN in York was generally higher than that in Simcoe.

Assessing the global integration score

The global integration scores calculated by different approaches were similar. We
performed all bootstrap procedures by simulating 500, 1000, 5000, 10000, and 40000
bootstrap samples. Although, increasing the number of simulations reduced the random

sampling error caused by the bootstrap procedure itself, the results only differed in the

third decimal place. Thus, we only reported the results by simulating 40000 bootstrap

samples. The weighted-average approach provided a slightly different estimate in some

cases and the narrowest 95% CI (Figures 5 and 6). Other researchers also reported a

narrower confidence interval when using a weighted approach [32]. Still, the 95% CIs

covered all scores by different methods. The only exception was the P4 global score

measuring the overall agreement between self-perceived and self-expected involvements

in York, where the weighted-average method provided a significantly larger estimate than

other methods. The global scores estimated by different bootstrap procedures were

identical to the standard one. This showed that the average integration score was a simple

and reliable estimate of the global score. The findings were consistent with the fact that a

sample mean was an unbiased estimate of the population mean. The 95% CIs in bootstrap

methods were slightly more precise. Overall, the sample mean was a robust estimate of

the global integration score.

Discussion

We have developed a method for quantifying the degree of integration for agencies in an

integrated service network. Using this method, mangers could identify the current gap in

service integration. We applied the method in measuring the CTN agencies. Non-

respondent agencies had lower P1 integration score, which indicated a poorer integration

as perceived and expected by the group. For some agencies, their scores (i.e. P1, P2, P3,

and P4 scores) varied largely across different perspectives of integration. As shown in

Figures 7 and 8, the spider plots were helpful for examining the gap of integration by different views. When an agency had 100% integration, the plot would show a complete 'diamond'. Any defect on the 'diamond' would indicate imperfect integration from some perspective. For example, in Simcoe, Agency 4 had much lower P1 and P2 scores than its other scores: below 45% versus above 80%. This suggested that the level of involvement perceived by all agencies including Agency 4 itself did not meet other agencies' expectation. This was an indication that the group might have a wrong expectation on the level of involvement required for Agency 4. Agency 14 had a much lower P2 score than other scores: 55% versus above 80%. The level of involvement perceived by Agency 14 met its own expectation but not others. This was an indication that Agency 14 might have an improper perception on the level of involvement it was contributing to the group. Agencies 17 had much lower P1, P2, and P3 scores than its P4 score: below 62% versus above 85%. This showed that the level of involvement perceived by Agency 17 only met its own expectation but not others, and was different from others' perception too. Agency 21 had much lower P1, P3, and P4 scores than its P2 score: below 59% versus above 82%. Although Agency 21's perception on its current involvement met others' expectation, it was different from what others had perceived. Our findings helped CTN managers to diagnose deeper problems of integration and potential barriers in integrating multiple services. By measuring the service integration over time, we could also evaluate the improvement in agency working relationships and promote further dialogue in achieving better integration.

Our measurement had some limitations. First, we only focused on measuring the degree of collaborative involvement in service integration. There were other types of integration [33], e.g. the functional integration, which we did not measure. Second, the measurement only captured the integration among the planning group. We acknowledged that integration achieved at the planning table did not always reflect the degree of integration in real practice, for example, among the frontline teams of workers. Third, there was potential respondent bias because filling out the measure by representative(s) is subject to proxy reporting bias. Representatives might not give the same information that others from the same agency would give. Halo effect and end-aversion were two other potential sources of bias in our results [34]. Finally, we were not able to evaluate the impact of non-respondents on the results.

Integration models generally require some formal mechanism, such as networks or committees of local agencies, to plan, organize, and deliver multiple services together. There are many barriers to integration because current health, educational, social, rehabilitative services, etc. are funded independently. Relationship, politics, communication, process, structure, and conflict are common problems for the failure of integration [35-38]. The gains from integration are often difficult to sustain and we need a tool to measure them. Despite the limitations, our method provides a valid way to conceptualize and quantify the service integration among agencies. During the time of measurement, CTN was undergoing initial planning to link resources and support, organize services, and create a new governance model. This partly explained the low

degree of integration that CTN agencies had achieved. By quantifying the state of

integration, our measurement helped CTN agencies visualize their agreement in the

process of integration and generated important discussions for their next stage of

planning. The unique value of our tool is that it provides a relative score on the degree to

which the actual involvement agrees with the expected involvement. In this regard, our

method greatly extends the HSIM by bridging the perceived and the expected

integrations. Compared with the Ahgren & Axelsson's method [23], our approach does

not adopt an assumption that higher levels of integration are better. Another difference

was that the Ahgren & Axelsson's method uses specific criteria to define levels of

integration (e.g. referrals, guidelines, chains of care, network managers, and pooled

resources). Although this might be a more objective approach, the measure could not

distinguish the differences in the degree of involvement that should be present in a well-

functioning network. In addition, our method produces simple and reliable global

integration scores that quantify the integration of an entire network as a whole, which is

hardly addressed in the current literature. Our method possesses a good content relevance

and coverage on measuring integration by including: the spectrum of services, the number

of providers, the score of integration, and the perspective of agencies. Our method can

also differentiate the degree of integration between similar networks, as shown in case of

the CTN Simcoe and York.

We have used the Partnership Self-Assessment Tool [39, 40] to examine the association

between our integration scores and components of a collaborative process – synergy,

leadership, efficiency, administration and management, resources, decision-making, benefits, drawbacks, and satisfaction. Our results showed that synergy was strongly associated with integration [41]. Other components including leadership, administration, decision-making, and satisfaction were also associated with integration. The findings demonstrated some extent of convergent validity because our integration measure was related to other variables of the same construct to which it should be related [34]. In an on-going analysis, we are examining the linkage between the degree of integration and network capacity that includes the average wait-time and the caseload. For future studies, we can apply the measurement in other similar service networks or in the same network for continuous evaluation. By repeating the measurement, investigators can determine integration patterns over time and examine the connection between integration and network outcomes longitudinally. We have found a potential influence of provider team integration on the quality of life of children with complex needs [42]. Future studies can also examine the similar relationship between different patient outcomes and the degree of integration using our measure.

Concluding Remarks

In this paper, we introduce a method for measuring the degree of integration for agencies in an integrated service network. Using the method, we measured the integration of the CTN agencies in Simcoe and York. In CTN Simcoe, agencies' P1, P2, P3, and P4 integration scores varied from 25% to 82%, 4% to 83%, 13% to 83%, and 4% to 96%, respectively. In CTN York, agencies' P1, P2, P3, and P4 integration scores varied from

27% to 73%, 27% to 81%, 9% to 82%, and 23% to 97%, respectively. Most agencies had

a score of less than 65%, a relatively low level of integration. The results revealed

existing problems in integrating CTN services. As measured by the agreement between

every other agency's perception and expectation, the overall integration of CTN in

Simcoe and York was 44% (95% CI: 39% - 49%) and 52% (95% CI: 48% - 56%),

respectively. The sensitivity analysis showed that the average integration score was a

reliable and robust estimate of the global integration score. The measurement provided

timely information for decision-making and improving the integration of CTN. The key

implication is that every integrated service network needs a valid measurement to

evaluate whether or not the collaborative process has been implemented as planned.

Measuring the service integration should be the first step in this evaluation. However,

quantitative methods available for measuring integration have been scarce in the

literature. Our method greatly extends the existing measures of integration by quantifying

the agreement between agencies' perceived and expected levels of involvement. Our

approach is unique such that we quantify the integration from four different perspectives

to identify deeper problems in integrating multiple services provided by autonomous

agencies. We showed that the proposed method possessed a good extent of validity and

could be applied in measuring other integrations in a similar setting.

Authors' contributions

CY conceived the study, designed the integration scores, performed all analyses, interpreted the results, and drafted and revised the manuscript. GB advised on important intellectual content and revised the manuscript. VSG revised the manuscript and contributed to the interpretation of results. JB revised the manuscript and contributed to the statistical analysis of the study. LT contributed to the statistical design, interpretation of results and revision of the manuscript. All authors have read and approved the final manuscript.

References

1. Canadian Institute for Health Information. National health expenditure trends 1975 to 2010. Canada: CIHI; 2010 Oct. [cited 2012 21 May]. Available from: https://secure.cihi.ca/free_products/nhex_trends_report_2011_en.pdf.

2. Canadian Institute for Health Information. Health Care in Canada 2010. Canada: CIHI; 2010 Dec. [cited 2012 21 May]. Available from: https://secure.cihi.ca/estore/productSeries.htm?pc=PCC64.

3. Skinner BJ, Rovere M. Paying more, getting less: Measuring the sustainability of government health spending in Canada 2009 Report. Canada: Fraser Institute; 2009 Nov. [cited 2012 21 May]. Available from: http://www.ihe.ca/publications/health-db/geo/448/.

4. The Conference Board of Canada. Understanding health care cost drivers and escalators. Ottawa; 2004 Mar. [cited 2012 21 May]. Available from: http://www.health.alberta.ca/documents/Health-Costs-Drivers-CBC-2004.pdf.

5. Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. The Lancet 2009;374(9696):1196-208.

6. Canadian Health Services Research Foundation. Myth: Canada's system of healthcare financing is unsustainable. Ottawa: CHSRF; 2007 Jan. [cited 2012 21 May]. Available from: http://28784.vws.magma.ca/mythbusters/html/myth28_e.php.

7. Stuart N, Adams J. The sustainability of Canada's healthcare system: A framework for advancing the debate. Healthcare Quarterly 2007;10(2):96-102.

8. Denton FT, Spencer BG. Chronic health conditions: Changing prevalence in an aging population and some implications for the delivery of health care services. Canadian Journal on Aging 2010;29(1):11-21.

9. Inkelas M, Raghavan R, Larson K, Kuo AA, Ortega AN. Unmet mental health need and access to services for children with special health care needs and their families. Ambulatory Pediatrics 2007;7(6):431-8.

10. Browne G, Roberts J, Gafni A, Byrne C, Kertyzia J, Loney P. Conceptualizing and validating the human services integration measure. International Journal of Integrated Care [serial online] 2004 May 19;4. [cited 2012 21 May]. Available from: http://www.ijic.org/index.php/ijic/article/view/98/195..

11. Topping S, Calloway M. Does resource scarcity create interorganizational coordination and formal service linkages? A case study of a rural mental health system. Advances in Health Care Management 2000;1(1):393-419.

12. Donaldson MS. Continuity of care: a reconceptualization. Medical Care Research and Review 2001;58(3):255-90.

13. Strandberg-Larsen M, Krasnik A. Measurement of integrated healthcare delivery: a systematic review of methods and future research directions. International Journal of Integrated Care [serial online] 2009 Feb 4;9. [cited 2012 21 May]. Available from: http://www.ijic.org/index.php/ijic/article/view/305/608.

14. Fleury M, Mercier C. Integrated local networks as a model for organizing mental health services. Administration and Policy in Mental Health 2002;30(1):55-73.

15. Bernabel R, Landl F, Ganbassi G, Sgadari A, Zuccala G, Mor V, et al. Randomised trial of impact of model of integrated care and case management for older people living in the community. British Medical Journal 1998;316(7141):1348-51.

16. Beland F, Bergman H, Lebel P, Clarfield AM, Tousignant P, Contandriopoulos AP, et al. A system of integrated care for older persons with disabilities in Canada: Results from a randomized controlled trial. Journal of Gerontology Series A: Biological Sciences and Medical Sciences 2006;61(4):367-73.

17. Larsen DL, Cannon W, Towner S. Longitudinal assessment of a diabetes care management system in an integrated health network. Journal of Managed Care Pharmacy 2003;9(6):552-8.

18. Browne G, Byrne C, Robert J, Gafni A, Whittaker S. When the bough breaks: Provider-initiated comprehensive care is more effective and less expensive for sole support parents on social assistance. Social Science and Medicine 2001;53(12):1697-710.

19. Browne G. More effective/less expensive health services also address the social determinants of health. Journal of the Ontario Association of Social Workers 2011 Mar 11;37(1). [cited 2012 21 May]. Available from: http://newsmagazine.oasw.org/magazine.cfm?magazineid=8&articleid=143.

20. Hebert RR, Veil AA. Monitoring the degree of implementation of an integrated delivery system. International Journal of Integrated Care [serial online] 2004 Sep 20;4. [cited 2012 20 Jul]. Available from: http://www.ijic.org/index.php/ijic/article/view/106/211.

21. Leonard BL, Parasuraman A. Marketing Services: Competing Through Quality. New York: Free Press; 1991. p. 20-56.

22. Wakefield DS, Halbesleben JR, Ward MM, Qiu Q, Brokel J, Crandall D. Development of a measure of clinical information systems expectations and experiences. Medical Care 2007;45(9):884-90.

23. Ahgren B, Axelsson R. Evaluating integrated health care: a model for measurement. International Journal of Integrated Care [serial online] 2005 Aug 31;5. [cited 2012 21 May]. Available from: http://www.ijic.org/index.php/ijic/article/view/134/267.

24. Browne G. Commentary: "Evaluating integrated health care: a system of measurement". International Journal of Integrated Care [serial online] 2005 Aug 31;5. [cited 2012 20 Jul]. Available from:
http://www.ijic.org/index.php/ijic/article/view/131/262.

25. Browne G, Kingston D, Grdisa V, Markle-Reid M. Conceptualization and measurement of integrated human service networks for evaluation. International Journal of Integrated Care [serial online] 2007 Dec 20;7. [cited 2012 20 Jul]. Available from:
http://www.ijic.org/index.php/ijic/article/view/URN%3ANBN%3ANL%3AUI%3A10
-1-100456/427. URN:NBN:NL:UI:10-1-100456.

26. Brazil K, Whelan T, O'Brien MA, Sussman J, Pyette N, Bainbridge D. Towards improving the co-ordination of supportive cancer care services in the community. Health Policy 2004;70(1):123-31.

27. Austin PC. Using the bootstrap to improve estimation and confidence interval for regression coefficients selected using backwards variable elimination. Statistics in Medicine 2008;27(17):3286-300.

28. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1993.

29. Shao J, Tu D. The jackknife and bootstrap. New York: Springer-Verlag; 1995.

30. Gleason JR. Algorithms for balanced bootstrap simulations. The American Statistician 1988;42(4):263-6.

31. Rubin DB. The Bayesian bootstrap. The Annals of Statistics 1981;9(1):130-4.

32. O'Gorman TW. Using adaptive weighted least squares to reduce the lengths of confidence intervals. The Canadian Journal of Statistics 2001;29(3):459-71.

33. Gillies RR, Shortell SM, Anderson DA, Mitchell JB, Morgan KL. Conceptualizing and measuring integration: findings from the Health Systems Integration Study. Hospital & Health Services Administration 1993;38(4):467-89.

34. Streiner DL, Geoff N. Health measurement scales: a practical guide to their development and use. 4th ed. New York: Oxford University Press; 2008.

35. Bodenheimer T. Coordinating care - A perilous journey through the health care system. New England Journal of Medicine 2008;358(19):1064-71.

36. Lawrence PR, Lorsch JW. Differentiation and integration in complex organizations. Administrative Science Quarterly 1967;12(1):1-47.

37. Sandfort JR. Exploring the Structuration of Technology within Human Service Organizations. Administration and Society 2003;34(6):605-31.

38. Sandfort JR. Why us human services integration so difficult to achieve? Focus 2004;23(2):35-9.

39. Center for the Advancement of Collaborative Strategies in Health. Partnership Self-Assessment Tool. New York; 2006 Jun. [cited 2012 17 May]. Available from: http://partnershiptool.net.

40. Weiss ES, Anderson RM, Lasker RD. Making the most of collaboration: Exploring the relationship between synergy & partnership functioning. Health Education & Behavior 2002;29(6):683-98.

41. Ye C. The Measurement of Integrated Human Services Network (The Children's Treatment Network of Simcoe York). Hamilton: McMaster University; 2008 Dec. [cited 2012 17 May]. Available from: http://digitalcommons.mcmaster.ca/opendissertations/4697.

42. Thurston S, Paul L, Ye C, Loney P, Browne D, Browne G, et al. System integration and its influence on the quality of life of children with complex needs. International Journal of Pediatrics 2010 Aug 11;2010. [cited 2012 15 May]. Available from: http://www.hindawi.com/journals/ijped/2010/570209/.

APPENDIX – Statistical Methods

Calculating the integration scores

The small letter *n* denotes the number of respondents. Since an agency does not rate it

itself, the number of respondents is the total number of agencies less 1 when there is no

non-respondent. The integration score $\hat{x}$ is calculated by

$$\hat{x} = \frac{\sum_{i=1}^{n} I_i(\text{perceved involvement=expected involvement})}{n},$$

where I_i is an indicator that returns 1 if the perceived involvement is equal to the expected

involvement and 0 otherwise for the i_{th} pair of ratings. Current perception of integration

can be either self- or group-perceived. Similarly, expectation can be either self- or group-

expected.

The variance of integration score is calculated by

$$\frac{\hat{x}(1-\hat{x})}{n},$$

where $\hat{x}$ is the P1, P2, P3 or P4 score.

Calculating the global integration score

The standard method

The capital letter *N* denotes the total number of agencies within a network. The average

integration score estimates the global integration $\hat{\theta}$, given by

$$\hat{\theta} = \frac{\sum_{i=1}^{N} \hat{x}_i}{N},$$

where $\hat{x}_i$ is the estimated P1, P2, P3, or P4 score of the i_{th} agency.

The 95% confidence interval (CI) of a global integration score is calculated by

$$(\hat{\theta}\text{-}1.96\,\frac{s}{\sqrt{n}}, \hat{\theta} + 1.96\,\frac{s}{\sqrt{n}}),$$

where s is the sample standard deviation of the integration scores.

The weighted-average method

The weighted global integration score is calculated by

$$\hat{\theta}_{\text{weighted}} = \frac{\sum_i^n \hat{x}_i w_i}{\sum_i^n w_i},$$

where $\hat{x}_i$ is the integration score of the i_{th} agency and w_i is the inverse of variance of $\hat{x}_i$.

The 95% CI of the weighted global integration score is calculated by

$$(\hat{\theta}\text{-}1.96\,\sqrt{\frac{1}{\sum_i^n w_i}}, \hat{\theta} + 1.96\,\sqrt{\frac{1}{\sum_i^n w_i}}).$$

The bootstrap method

In the standard bootstrap, we randomly draw an integration score from all agencies with replacement repeatedly to generate a bootstrap sample. We repeat that procedure to generate a large number of *k* bootstrap samples. The choice of *k* often depends on the computational power. In the balanced bootstrap, every integration score appears the same number of times among bootstrap samples. In other words, each original integration score is selected *k* times and allocated randomly in the *k* bootstrap samples. We achieve that by creating *k* replicates of original sample and then randomly sampling without replacement among those replicates to create *k* equal bootstrap samples. We calculate the mean of each bootstrap sample by

$$\hat{\theta}_{b_j} = \frac{\sum_{i=1}^N \hat{x}_i^b}{N}, j = 1,2,\dots k,$$

where $\widehat{\theta_{b_j}}$ is the average integration score for the j_{th} bootstrap sample and $\hat{x}_i^b$'s are the

integration scores selected in that bootstrap sample.

When k is large enough (usually over 1000), aggregating the means of all bootstrap

samples constitutes a sampling distribution of the global score. The bias in estimating the

global score can be then approximated by

$$\widehat{\text{Bias}} = \frac{\sum_{j=1}^{k} \widehat{\theta}_{b_j}}{k} - \hat{\theta} = \hat{\theta}_b - \hat{\theta},$$

where $\widehat{\theta_b}$ is the expected global integration score from bootstrap samples and $\hat{\theta}$ is the

mean of the original sample.

The bias-adjusted bootstrap estimate of the global integration score is calculated by

$$\hat{\theta}_{\text{bias-adjusted}} = \hat{\theta} - \widehat{\text{Bias}} = 2\hat{\theta} - \hat{\theta}_b.$$

The 95% bias-corrected and accelerated (BCa) confidence interval for the global

integration score is (m_1, m_2) where

$$m_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 - 1.96}{1 - \hat{a}(\hat{z}_0 - 1.96)}\right) \text{ and } m_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + 1.96}{1 - \hat{a}(\hat{z}_0 + 1.96)}\right).$$

The $\Phi(\cdot)$ is the standard normal cumulative function; $\hat{a}$ $z^{(\alpha)}$ is the acceleration; and $\hat{z}_0$ is

the bias-correction. The method of calculating $\hat{a}$ and $\hat{z}_0$ empirically is provided in [28].

In the Bayesian bootstrap, each bootstrap sample generates a posterior probability for

each integration score in the original sample. We draw N -1 uniform random numbers

between 0 and 1, order them to be $c_1, c_2, ..., c_N$, and let $c_0 = 0$ and $c_1 = 1$. We then

calculate the gaps $g_i = c_i - c_{i-1}, i = 1, ..., N$ and attach $g = (g_1, ..., g_N)$ as the vector of

probabilities to the original integration scores $\hat{x}_1, ..., \hat{x}_N$ for that bootstrap sample. We

repeat that procedure to create k bootstrap samples. Details of the procedure can be found

in [31]. The posterior estimate of the average integration scores in each bootstrap sample

is calculated by,

$$\hat{\theta}_{b_j} = \sum_i^N g_i \hat{x}_i \ , j = 1, 2, \dots, k.$$

The Bayesian bootstrap (BB) estimate of the global integration score can be calculated by

the average of all posterior estimates from bootstrap samples, i.e.

$$\hat{\theta}_{BB} \frac{\sum_{j=1}^k \hat{\theta}_{b_j}}{k}.$$

The 95% credibility interval can be obtained from the empirical distribution of all

posterior estimates $\hat{\theta}_{b_j}$'s.

Figure 1: The Human Service Integration Measure Scale

To what extent **are you** involved with the following services?		AGENCY/		To what extent **should you** be involved with the following services?
Rating Scale	Rate (0-4)	Do not rate your own service	Rate (0-4)	**Rating Scale**
0 = No awareness: Your agency **is not** aware of the other service		Agency A		**0 = No awareness:** Your agency **should not** be aware of the other service
		Agency B		
1 = Awareness: You **have** knowledge of the other service although no effort is taken to organize activities according to any principles except those that conform to individual agency missions.		Agency C		**1 = Awareness:** You **should** have knowledge of the other service although no effort is taken to organize activities according to any principles except those that conform to individual agency missions.
		Agency D		
		Agency E		
		Agency F		
2 = Communication: You and the other service **have** an active program of communication and information sharing.		Agency G		**2 = Communication:** You and the other service **should** have an active program of communication and information sharing.
		Agency H		
		Agency I		
		Agency J		
3 = Cooperation: You and the other service **each use** your knowledge of the other's service to guide and modify your own service planning in order to obtain a better set of links between services.				**3 = Cooperation:** You and the other service **should** each use your knowledge of the other's service to guide and modify your own service planning in order to obtain a better set of links between services.
4 = Collaboration: You and the other service **jointly** plan the offering of service and actively modify service activity based on advice and input from mutual discussions.				**4 = Collaboration:** You and the other service **should** jointly plan the offering of service and actively modify service activity based on advice and input from mutual discussions.

Figure 2: An example for organizing the data

AGENCY A		
Perceived	Do not rate your own service	Expected
X	Agency A	X
1	Agency B	1
1	Agency C	2
1	Agency D	3

AGENCY B		
Perceived	Do not rate your own service	Expected
2	Agency A	1
X	Agency B	X
2	Agency C	2
2	Agency D	3

AGENCY C		
Perceived	Do not rate your own service	Expected
3	Agency A	1
3	Agency B	2
X	Agency C	X
3	Agency D	3

AGENCY D		
Perceived	Do not rate your own service	Expected
4	Agency A	1
4	Agency B	2
4	Agency C	3
X	Agency D	X

Perceived involvement – square matrix

	Agency A	Agency B	Agency C	Agency D
Agency A	X	2	3	4
Agency B	1	X	3	4
Agency C	1	2	X	4
Agency D	1	2	3	X

Expected involvement – square matrix

	Agency A	Agency B	Agency C	Agency D
Agency A	X	1	1	1
Agency B	1	X	2	2
Agency C	2	2	X	3
Agency D	3	3	3	X

	Self-perceived			Self-expected			Group-perceived			Group-expected		
A	1	1	1	1	2	3	2	3	4	1	1	1
B	2	2	2	1	2	3	1	3	4	1	2	3
C	3	3	3	1	2	3	1	2	4	2	2	3
D	4	4	4	1	2	3	1	2	3	3	3	3

Figure 3: The framework of measuring integration
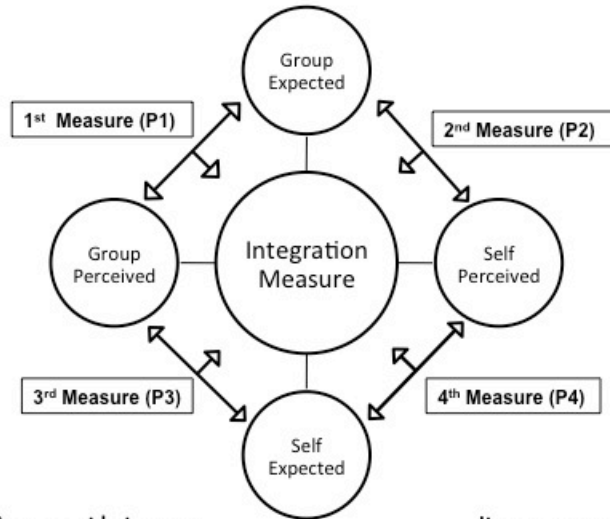
Integration Score	Agreement between...	It measures...
P1	Group-perceived rating and group-expected rating	The degree of agreement between an agency's involvement perceived by others and its involvement expected by others.
P2	Self-perceived rating and group-expected rating	The degree of agreement between an agency's self-perceived involvement and its involvement expected by others.
P3	Group-perceived rating and self-expected rating	The degree of agreement between an agency's involvement perceived by others and the involvement expected itself.
P4	Self-perceived rating and self-expected rating	The degree of agreement between an agency's self-perceived involvement and the involvement expected itself.

Figure 4: An example of the global integration

Figure 5: Global integration scores of CTN Simcoe

Figure 6: Global integration scores of CTN York

Figure 7: The spider plot for comparing the 4 integration scores for agencies in Simcoe

Figure 8: The spider plot for comparing the 4 integration scores for agencies in York

Table 1: Integration scores for the agencies of the Children's Treatment Network

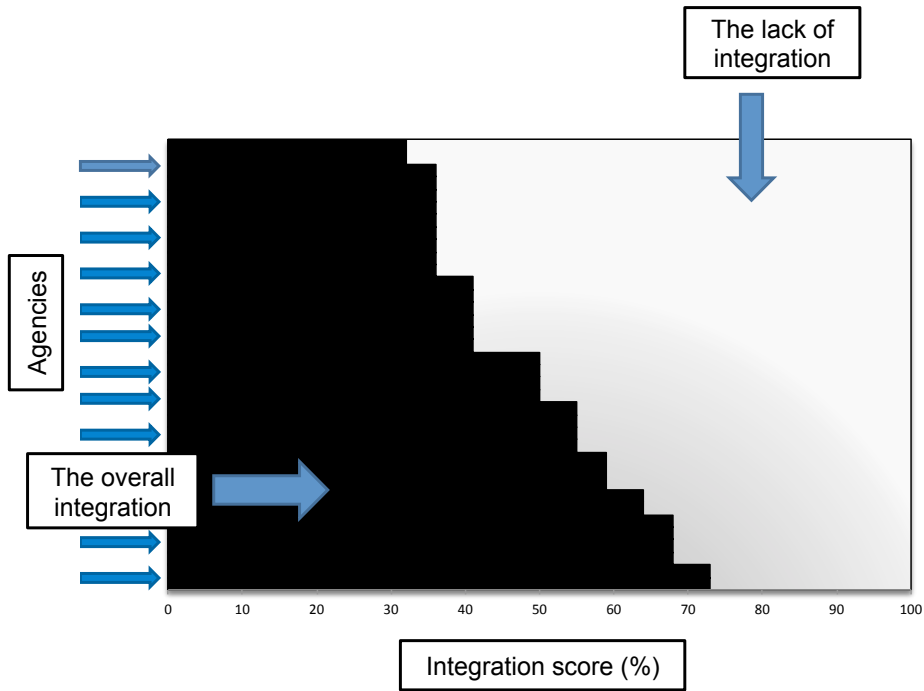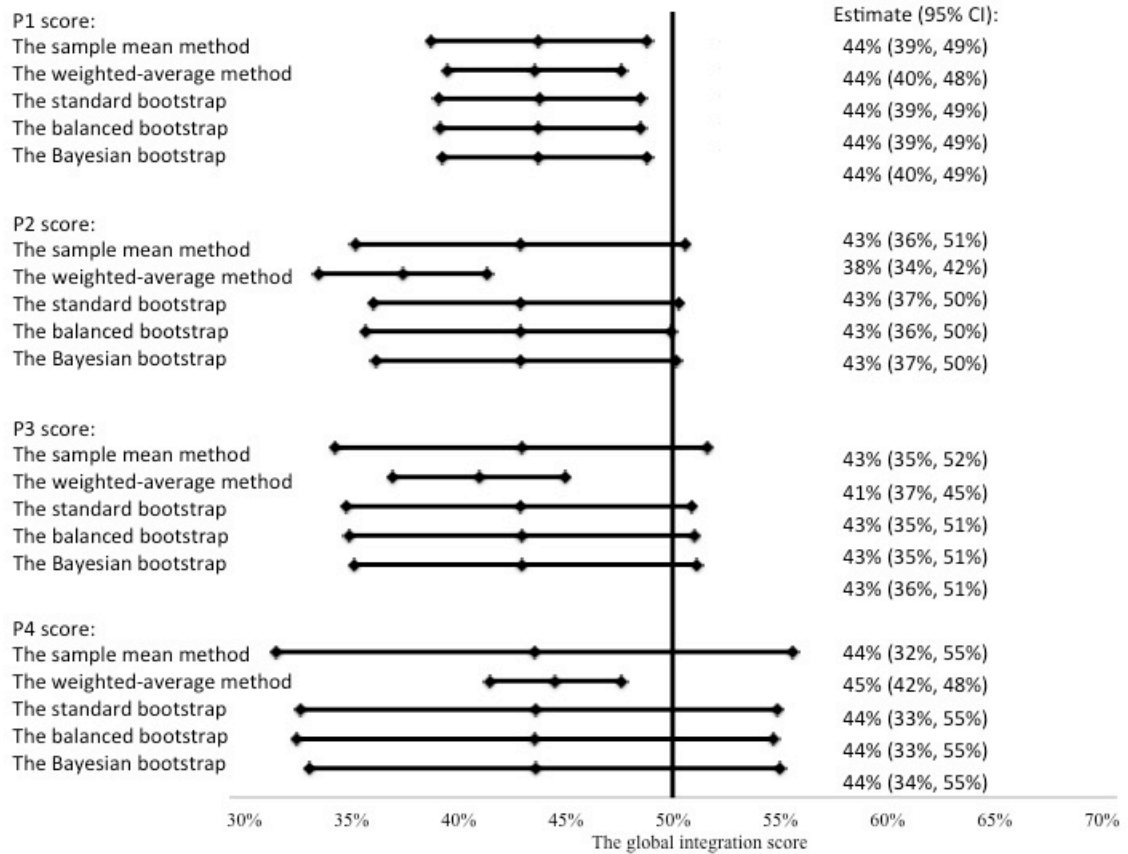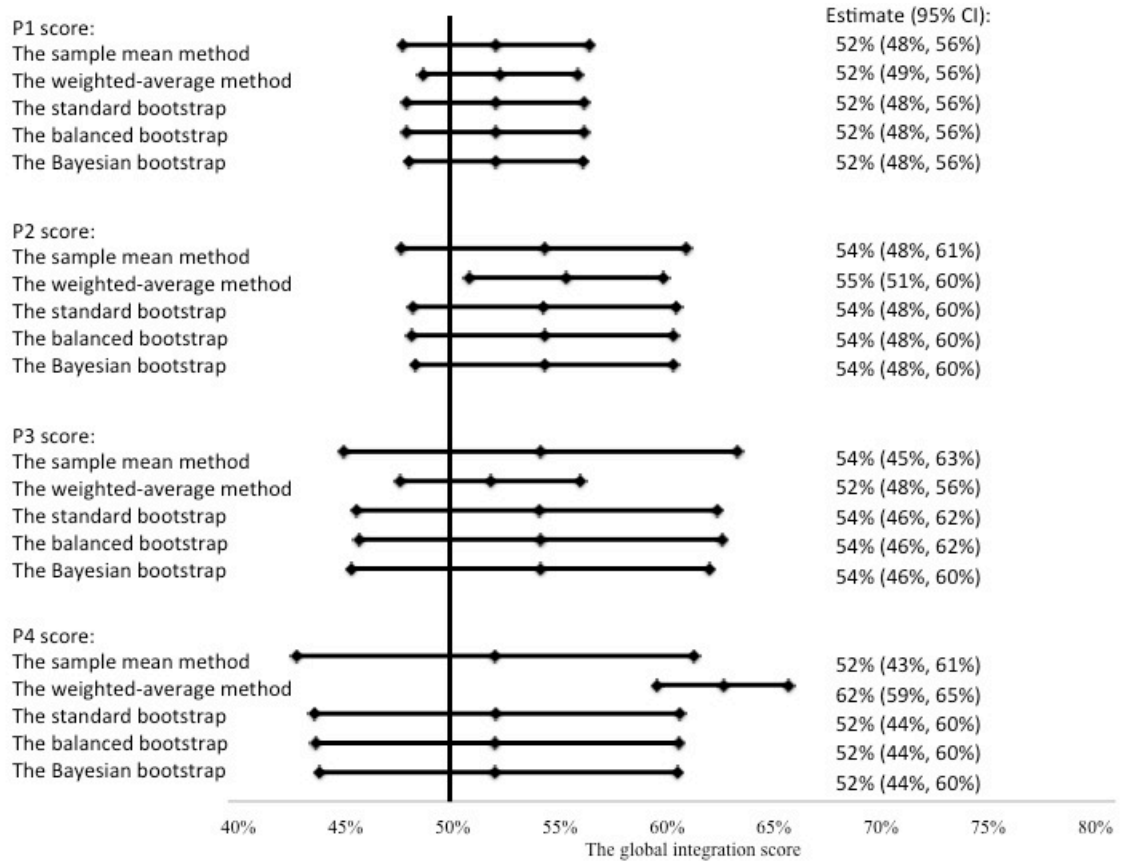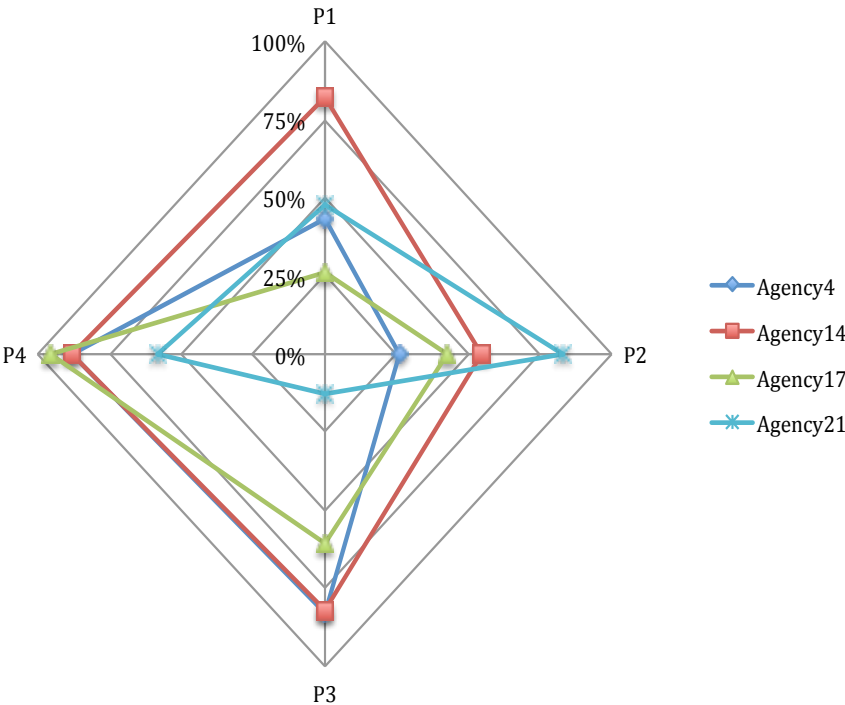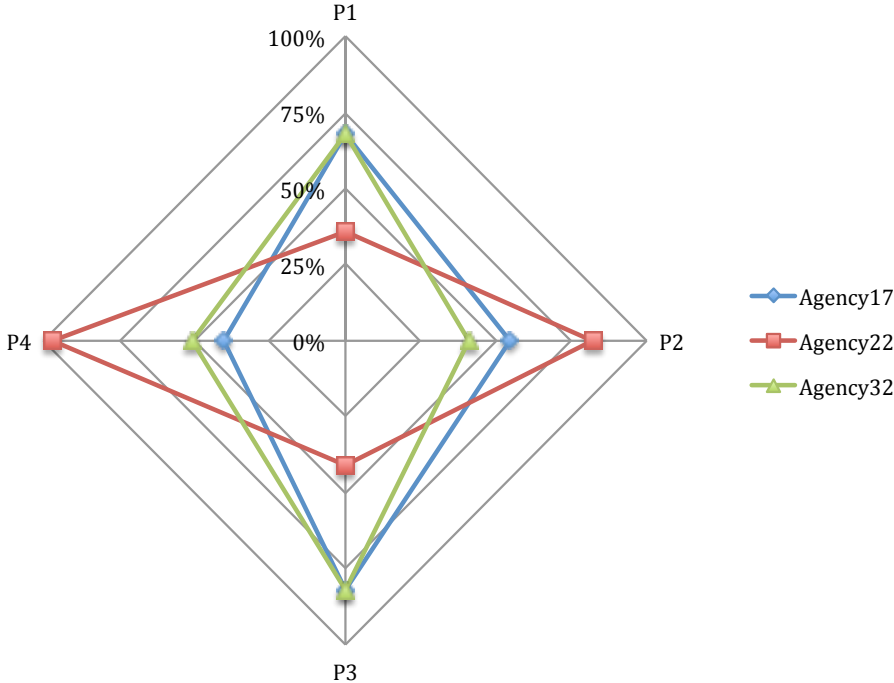| Service agency	Simcoe (n=27)				York (n=36)			
	P1	P2	P3	P4	P1	P2	P3	P4
1	30%	39%	26%	12%	41%	62%	14%	24%
2	38%	-	-	-	68%	59%	59%	46%
3	43%	30%	35%	31%	35%	-	-	-
4	43%	26%	83%	88%	41%	50%	27%	43%
5	39%	43%	26%	23%	52%	-	-	-
6	61%	65%	35%	54%	68%	32%	64%	26%
7	35%	35%	65%	38%	52%	-	-	-
8	25%	-	-	-	61%	-	-	-
9	25%	-	-	-	64%	36%	45%	31%
10	48%	26%	48%	50%	73%	68%	82%	79%
11	39%	35%	22%	12%	57%	-	-	-
12	43%	30%	43%	31%	59%	50%	64%	74%
13	26%	35%	22%	19%	64%	59%	68%	49%
14	82%	55%	82%	88%	50%	64%	59%	74%
15	57%	39%	65%	38%	43%	-	-	-
16	52%	45%	41%	12%	43%	-	-	-
17	26%	43%	61%	96%	68%	55%	82%	40%
18	48%	61%	43%	85%	55%	45%	73%	49%
19	48%	48%	57%	65%	64%	-	-	-
20	48%	4%	57%	4%	41%	55%	50%	57%
21	48%	83%	13%	58%	32%	36%	9%	24%
22	48%	48%	52%	42%	36%	82%	41%	97%
23	43%	74%	13%	50%	41%	48%	62%	68%
24	52%	35%	39%	15%	43%	-	-	-
25	52%	70%	22%	19%	59%	73%	45%	49%
26	48%	35%	30%	69%	62%	71%	52%	51%
27	39%	30%	57%	50%	27%	-	-	-
28					57%	-	-	-
29					55%	27%	50%	23%
30					43%	-	-	-
31					36%	50%	27%	31%
32					68%	41%	82%	51%
33					73%	68%	77%	77%
34					57%	-	-	-
35					41%	41%	64%	66%
36					48%	76%	48%	69%

*Percentages were rounded up to integer; the agencies were listed in a consecutive order and the same number **did not** refer to the same agency; CI = confidence interval.

P1= the degree of agreement between an agency's involvement perceived by others and its involvement expected by others
P2= the degree of agreement between an agency's self-perceived involvement and its involvement expected by others
P3= the degree of agreement between an agency's involvement perceived by others and the involvement expected itself
P4= the degree of agreement between an agency's self-perceived involvement and the involvement expected itself

Table 2: Global integration scores estimated by different methods

Region	Method	P1	95% CI*	P2	95% CI*	P3	95% CI*	P4	95% CI*
	Sample mean	44%	(39%, 49%)	43%	(36%, 51%)	43%	(35%, 52%)	44%	(32%, 55%)
	Other methods								
Simcoe	Weighted average	44%	(40%, 48%)	38%	(34%, 42%)	41%	(37%, 45%)	45%	(42%, 48%)
	Standard bootstrap	44%	(39%, 49%)	43%	(37%, 50%)	43%	(35%, 51%)	44%	(33%, 55%)
	Balanced bootstrap	44%	(39%, 49%)	43%	(36%, 50%)	43%	(35%, 51%)	44%	(33%, 55%)
	Bayesian bootstrap	44%	(40%, 49%)	43%	(37%, 50%)	43%	(36%, 51%)	44%	(34%, 55%)
	Sample mean	52%	(48%, 56%)	54%	(48%, 61%)	54%	(45%, 63%)	52%	(43%, 61%)
	Other methods								
York	Weighted average	52%	(49%, 56%)	55%	(51%, 60%)	52%	(48%, 56%)	62%	(59%, 65%)
	Standard bootstrap	52%	(48%, 56%)	54%	(48%, 60%)	54%	(46%, 62%)	52%	(44%, 60%)
	Balanced bootstrap	52%	(48%, 56%)	54%	(48%, 60%)	54%	(46%, 62%)	52%	(44%, 60%)
	Bayesian bootstrap	52%	(48%, 56%)	54%	(48%, 60%)	54%	(46%, 62%)	52%	(44%, 60%)

*CI = confidence interval, a credibility interval was calculated instead in the Bayesian bootstrap method; n = number of respondents; bootstrap estimates were obtained by simulating 40000 bootstrap samples.

P1= the degree of agreement between an agency's involvement perceived by others and its involvement expected by others

P2= the degree of agreement between an agency's self-perceived involvement and its involvement expected by others

P3= the degree of agreement between an agency's involvement perceived by others and the involvement expected itself

P4= the degree of agreement between an agency's self-perceived involvement and the involvement expected itself

CHAPTER 3

A sensitivity analysis of the Children's Treatment Network trial: a randomized controlled trial of integrated services versus usual care for children with special health care needs

Chenglin Ye[1,2], Gina Browne[1,3], Joseph Beyene[1], Lehana Thabane[1,2*]

Affiliations:

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton ON, L8S 4L8 Canada

[2]Biostatistics Unit, St. Joseph's Healthcare Hamilton

[3]School of Nursing, McMaster University, Hamilton ON, L8S 4L8 Canada

[*]Corresponding author:

Lehana Thabane

Email: thabanl@mcmaster.ca

Phone: +1/905-522-1155 ext.33720

Fax: +1/905-308-7212

Cite as: Ye C, Browne G, Beyene J, Thabane L. A sensitivity analysis of the Children's Treatment Network trial: a randomized controlled trial of integrated services versus usual care for children with special health care needs. Clinical Epidemiology 2013;5:373-385.

Abstract

Background: The value of integrated care through comprehensive, coordinated, and family-centered services has been increasingly recognized for improving health outcomes of children with special health care needs (CSHCN). In a randomized controlled trial (RCT), the integrated care provided through the Children's Treatment Network (CTN) was compared with usual care on improving psychosocial health of the target CSHCN. In this paper, we aimed to estimate the effect of the CTN care by conducting multiple analyses to handle non-compliance in the trial.

Methods: The trial recruited the target children in Simcoe County and York Region, Ontario Canada. Children were randomized to receive the CTN or usual care and were followed for 2 years. The CTN group received integrated services through multiple providers to address their specific needs while the usual care group continued to receive care directed by their parents. The outcome was the change of psychosocial quality of life (QOL) in 2 years. We conducted the intention-to-treat, as-treated, per-protocol, and instrumental variable analyses to analyze the outcome..

Results: The trial randomized 445 children, with 229 in the intervention group and 216 in the control group. During follow-up, 52% of the children in the intervention group did not receive complete CTN care for various reasons. At 2 years, we did not find significant improvement in psychosocial QOL among the children receiving the CTN care compared with usual care (intention-to-treat mean difference: 1.50, 95% confidence interval: -1.49, 4.50; p=0.32). Other methods of analysis yielded similar results.

Conclusion: Although the effect of the CTN care was not significant, there was evidence which showed the benefits of integrated care for CSHCN. More RCTs are needed to demonstrate the magnitude of such effect. The CTN study highlights the key challenges in RCTs to assess interventions of integrated care. It informs further RCTs of similar evaluations.

Keywords: CSHCN, chronically ill, coordinated, family-centered care, RCT, non-compliance.

Introduction

The "children with special health care needs" (CSHCN) are those "who have or are at increased risk for a chronic physical, developmental, behavioral, or emotional condition, and who also require health and related services of a type or amount beyond that required by children generally".[1] In Canada, about 13% to 18% of children are considered to have special needs.[2] These children usually have poor quality of life[3] associated with emotional symptoms,[4] psychological problems,[5] and health risk behaviours.[6] Two thirds of parents with disabled child do not get necessary treatments for their child and more than one third of them do not know where to look for help.[7] These parents endure enormous financial and mental burdens due to their child's illnesses.[8] Since most CSHCN live at home, the parents have to quit their jobs, work fewer hours, turn down a promotion, or compromise job performance[9] for taking care of their child with special needs. "Finding services, battling waitlists, travelling long distances, briefing numerous professionals from different sectors on their child's history, and coordinating services from multiple agencies" are just part of their daily life.[10] These parents commonly describe continuing strain and chronic fatigue concerning their child's well-being[11] and over 35% of them report high levels of emotional distress.[12] Their families are also affected by high divorce rates.[13]

CSHCN are usually diverse in nature and need an individualized approach to address their specialized and multiple needs.[14] The value of continuous, comprehensive, coordinated, and family-centered care[15] has been widely recognized. Delivery of this new care requires a more integrative approach than the traditional one through: collaboration

with parents to decide case-specific goals; coordination of an individual team of service providers; and provision of a single point of access to address families' multiple needs. As an example, medical home[16] in the United States is a new development on health care delivery for the families with CSHCN. It has led to reduction in wait-time or unmet health care needs for the target children;[17] increased use of preventive care;[18] decreased risk for comorbidities;[19] and alleviation of family's burdens.[20] Adoption of those strategies in Canadian system has been limited. There are examples of innovative interventions (for example, the Chronic Care Model in British Columbia and the Family Health Network in Ontario) but these interventions target adult patients.[14] Compared with adults, CSHCN are affected by a larger number of uncommon conditions and demand more tailored services and a more complex model of care.

In 2006, a new model of care for CSHCN and their families, the Children's Treatment Network (CTN), was launched in Simcoe County and York Region in Ontario Canada. The CTN model of care is funded by the government and is based on the collaboration of local service providers from different agencies and organizations. It provides a single point of access and service navigation for the children. The CTN model is unique such that it assigns a service coordinator and an individual team of health service providers for each target family; develops a single plan of care; and uses an electronic record system to share clinical information and assessment of the child. Some services available to families through the CTN include developmentally appropriate therapy, speech therapy, augmentative communication, in-home social support, behaviour therapy, and early childhood educator (www.ctn-simcoeyork.ca). Before the launch of the

CTN, parents in both communities had limited access to specialized treatments needed for their children and had to travel outside of the region to get those services. Funded by the Ministry of Children and Youth Services and the Ministry of Health and Long-term Care In Ontario Canada, Thurston, Paul, and Browne et al.[21] conducted a randomized controlled trial to compare the integrated care interventional program organized through the CTN versus usual care for managing children with multiple and complex health needs in Simcoe County and York Region (trial registration in clinicaltrials.gov: NCT01379443). This CTN trial was subject to substantial non-compliance where about 52% of the children in the intervention group only received parts of the CTN integrated care. Some analyses of the CTN were published previously.[22,23]

Objectives

The primary goal of the CTN trial was to assess the effect of CTN integrated care on improving the psychosocial function of the target children in Simcoe County and York Region. In this study, we compared the intention-to-treat (ITT) with alternative methods to analyze the data. We also aimed to assess the sensitivity of the ITT analysis under participants' non-compliance to the CTN intervention.

Material and methods

We describe the CTN trial here by adopting the Consolidated Standards of Reporting Trials (CONSORT) guideline.[24] Using a parallel design, each eligible child was randomized to the intervention or control group with a 1:1 allocation ratio.

Participants

Participants were recruited from the Simcoe County and York Region in Ontario Canada. The Simcoe York District Health Council[25] projected the total number of children in both regions to be 390,498 in 2007. According to the national survey, 1% of children were estimated to have severe disabilities and receive rehabilitation services.[26]

Children (0 to 19 years old) were recruited if they had any of the following conditions: cerebral palsy, brain injury, developmental difficulties, down syndrome, spina bifida, autism, physical disability, developmental disability, pervasive developmental disorder, and chronic medical conditions. Children who were receiving palliative care, requiring emergency services, or living outside of the region at the time were excluded. Non-English speaking families without an English translator were also excluded.

Potential children were identified through the Community Care Access Centres, the school boards, and the Early Intervention Services agencies. Eligible families were contacted by mail with an invitation package, which contained a study information letter, parental consent and participant contact information form. A second mail-out was done to families who did not respond to the initial mail-out. Only the children whose parents had provided written informed consent could participate. Children were enrolled from May to December 2007. A unique patient ID was created for each enrolled child.

Ethics

There were no known harms or safety risks to the children and their families involved in the trial. The research ethics board at Hamilton Health Sciences/McMaster Health Sciences approved the study.

Intervention

Each child in the CTN group was assigned a service navigator who conducted a comprehensive assessment to identify the child's health conditions. A trained service coordinator then followed up with the family. In working with the parents, an individual team of service providers according to the child's health and social needs was formed. For example, a team of augmentative communication services consultant, early interventionist, occupational therapist, physiotherapist, speech and language therapist, and service coordinator was assembled for a child with cerebral palsy. This integrative team together with family members developed a single plan of care for the child. The service coordinator organized the delivery of services according to the plan. On a regular basis, the team met with the family for on-going assessment and revision of the plan. All assessments and clinical notes on the child were documented in an electronic record system shared by all team members. Figure 1 shows the detailed components of the CTN integrated care. Families in the control group continued to manage services for their child in a self-directed manner. This group did not have access to the CTN integrated care; however, they had access to all other service providers (including CTN partners) as they requested.

The fidelity to integrated child and family team care was evaluated by: 1) content in the child and family team of providers (extent, scope, and agreement in collaboration); 2) the quality of team functioning measured by the Partnership Functioning Scale;[27] and the frequency and duration coverage of the prescribed services as outlined in the single plan of care compared with providers' entries on electronic record.

Outcome and variables

The outcome was the change of child's psychosocial quality of life (QOL) in 24 months. Child's psychosocial QOL was reported by parents using the short form Pediatric Quality of Life (SF PedsQL).[28] The psychosocial QOL was the sum of the emotional, social, and school functions and was extrapolated to have a score from 0 (the worst QOL) to 100 (the best QOL). Children's age, gender, and admission diagnosis were collected. The age, marital status, educational level, and family income of the parent who was the most knowledgeable (PMK) about the child were recorded.

Parental distress score:

The Kessler scale (K10)[29] was used to measure PMK symptoms of depression anxiety. Ten questions measured the feelings: sad, nervous, restless, hopeless, worthless, everything was an effort, tired for no good reason, so nervous that nothing could calm down, so restless and could not sit still, and depressed during the past month. PMK rated each item on a five-point scale (1 – "all of the time" to 5 – "none of the time"). Scores range from 10 (no symptom of distress) to 50 (severe distress).

Positive parenting score:

The National Longitudinal Survey of Children and Youth (NLSCY) parenting scale was used and 5 questions adapted from the Parent Practices Scale[30] were used to measure positive parenting behaviours (praise, play). PMK rated each item (e.g., "Do something special with your child that he/she enjoys") on a five-point scale (0 – "never" to 4 – "many times each day"). Higher scores indicate more positive parenting behaviours.

Social support score:

An eight item shortened version of the Social Provisions Scale[31] was used to measure the level of social support that the PMK received in guidance, reliable alliance, and attachment. PMK rated each item on a four-point scale (0 – "strongly disagree" to 3 – "strongly agree"). Higher scores represent greater level of social support.

Family function score:

Thirteen items taken from the NLSCY population survey, based on a subscale of the McMaster Assessment Device[32], were used to measure various aspects of family functioning in problem solving, communication, roles, affective responsiveness, affective involvement, and behaviour control. PMK rated each item on a four-point scale (0 – "strongly agree" to 3 – "strongly disagree"). Scores of negatively oriented items (e.g., "We avoid discussing our fears or concerns") were reversed so that higher scores represent greater family dysfunction.

Sample size

Using a minimum clinically important difference of 10 on the target children's psychosocial QOL[33] and an estimated standard deviation of 28 estimated from our pilot

sample, the sample size was calculated to be 240 for a two-sided test with 5% alpha and 80% statistical power. To allow for 10 independent variables and an attrition rate of 20%, the required sample size was 425.

Randomization

Using a parallel design, eligible children with parental consents were randomized to intervention and control groups with a 1:1 allocation ratio. Children were stratified by region (Simcoe/York), Community Care Access Centre (CCAC), and age (pre-school/school). They were randomized within stratum by using a block size of 6. The randomization list was generated by the Health and Social Service Utilization Research Unit (HSSURU) at McMaster University (fhs.mcmaster.ca/slru). The allocation codes were then sequentially linked to the patient IDs of enrolled children.

Blinding

Trained interviewers at HSSURU, who were blinded to group allocation, collected data using a standardized questionnaire booklet containing questions of demographics and outcome measures. The data collection was performed by telephone at baseline, 12, and 24 months. After enrollment, children were kept anonymously and were identified by their patient IDs. Only the statistician responsible for randomization had access to the allocation codes. The data analyst was blinded to group allocation; however, the participants were aware of their allocation. All data were stored and maintained at the HSSURU.

Statistical methods

We adopted the intention-to-treat (ITT) principle by analyzing all patients as how they were randomized regardless of whether they complied with the assigned treatment or not. The ITT estimate was obtained by fitting a linear regression model with adjustment for the baseline variables. Multiple imputation (MI) technique was used to impute the missing data by assuming that they were missing at random (MAR). Five imputation datasets were produced and combined to produce an estimate of 98.5% relative efficiency, given a fraction of about 30% missing data.[34]

In the sensitivity analysis of the impact of non-compliance, we employed three other analytical approaches: the as-treated (AT), the per-protocol (PP), and the instrumental variable (IV) analyses. In the AT analysis, we compared the treated children (those who received complete CTN integrated care) with the untreated ones (those who received parts or none of the CTN integrated care) to estimate the effect of the intervention. In the PP analysis, we analyzed the children by their randomization but only among the ones who complied with the assigned treatment. Children in the control group did not have access to the CTN integrated care and were deemed to comply with the usual care. Both AT and PP analyses compromised original randomization and were prone to selection bias and confounding. We employed propensity score (PS) technique to adjust for those factors. The application of PS in the AT and PP analyses could potentially be a novel approach. Since we could not find any study comparing different uses of PS in this context, we included all four commonly used PS methods: the matching, stratification, weighting, and

covariate adjustment. The details of those PS methods are given in the supplementary appendix.

The IV analysis[35] estimated the effect of the CTN integrated care through an IV. The randomization allocation (Z) was associated with the exposure of the CTN integrated care (X) but independent of the confounders and any factor that explained the outcome (Y). Thus, Z satisfied the conditions to be an IV and the treatment effect can be estimated by

$$\beta_{IV} = \frac{\beta_{(Z \to Y)}}{\beta_{(Z \to X)}},$$

where $\beta_{(Z \to Y)}$ is the association between randomization and the outcome; and $\beta_{(Z \to X)}$ is the association between randomization and the receipt of the CTN integrated care. We assumed that the Z-Y association was independent of the Z-X association. Under the assumption, we used the delta method to derive the variance of β_{IV} for calculating the 95% CI and p value. The details are provided in the supplementary appendix.

We compared the ITT with other methods of analysis to assess the sensitivity of the results. Figure 2 summarizes the different approaches of analysis. The results were expressed in mean difference (MD), 95% confidence interval (CI) of the MD, and corresponding p value. The level of significance was set at 5%. All analyses were performed in the statistical software R version 2.12.1 and SPSS IBM version 19.0.

Results

Recruitment, baseline data and participant flow

Between May and December 2007, 2319 eligible children were identified and their families were contacted by mail. Parents of 465 eligible children returned signed consent

forms to indicate their willingness to participate. Later, 20 consented parents did not complete the baseline interview and were excluded from the study. The remaining 445 children were randomized to either the intervention group (n = 229) or the control group (n = 216).

The children were followed until December 2009. During the 2-year follow-up, 53 children in the CTN group did not have a team of service providers assembled or did not have the services available for their specific needs. Another 58 children in the CTN group withdrew from the CTN integrated care. Common reasons for withdrawal were: child's decease, family's move outside of the region, parent's decision to manage child by herself, and transfer to another health care centre. Overall, 64 (28%) children in the intervention group and 57 (26%) children in the control group were lost to follow-up. Figure 3 shows the flow of participants in the study. All 445 children were analyzed by using multiple imputation technique.

Statistical analysis

The children's baseline characteristics are summarized in Table 1. Categorical and continuous variables are expressed in frequency (percentage) and mean (standard deviation), respectively. Within the CTN arm, the baseline variables were compared between the children who received complete CTN integrated care (treated) and those who received partial or none of the CTN integrated care (untreated). The results of the comparison are reported in Table 2. From the comparison, we did not find significant differences between treated and untreated children except for the parenting style (MD:

1.4; p=0.02) but a mean difference of 1.4 on a score ranging from 0 to 20 did not seem to be a clinical relevant association.

We conducted different analyses to estimate the effect of the CTN integrated care. The results did not show a significant improvement of the children's psychosocial QOL in the CTN group compared with the usual care group (MD: 1.50, 95% CI: -1.49,4.50; p=0.32). The results of estimated treatment effect are reported in Table 3. In general, the conclusion of the ITT analysis was consistent with those drawn from the AT, PP, and IV analyses. We also noticed some systematic patterns in the direction, magnitude, and precision of the estimates. All estimates except for the AT estimates favored the CTN integrated care. The PP estimates were the closest to 0 which represented no difference between the CTN integrated care and usual care. Both ITT and IV estimates showed a larger effect of the CTN integrated care over usual care than the PP estimates. The IV estimate had the largest 95% CI. For the AT and PP analyses which compromised original randomization, adjusting for confounders and imbalance by matching on the PS produced estimates of a wider 95% CI than did the other PS methods. Figure 4 shows a comparison of the estimates obtained from different analyses.

Discussion

The ITT analysis did not show a significant improvement of the psychosocial QOL among the children receiving the CTN integrated care than those receiving the usual care in Simcoe County and York Region over 2 years. We conducted AT, PP, and IV analyses to assess the sensitivity of this conclusion under substantial non-compliance to the CTN

intervention. Those alternative analyses also showed no significant difference between groups. Previously, two CTN studies were published[22,23], which explored the interactions between multiple factors and system integration on the child's psychosocial QOL and examined the associations between multiple factors and level of psychiatric distress experienced by the parents.

Our finding was limited by a number of factors. First, child's psychosocial QOL was reported by their parents. Parents might have limited knowledge concerning their children's health related QOL. The parents' responses reflected their own perception on children's disease-related experiences, which might not be the same as how the children felt. For example, it was found that children with a congenital below-the-elbow deficiency reported better QOL than what their parents perceived.[36] Second, the early development of the CTN model of care was associated with suboptimal intervention fidelity. The implementation of the CTN was a major undertaking in changing system and operations management. It took longer than anticipated for the network hosts to hire and train appropriate staff at the local team level, get local team sites operational, and engage all the needed agency partners' staff in the teams. The delay in creating the child and family teams was the key challenge in organizing services to some children. About 52% of the children assigned in the CTN group did not receive the CTN integrated care as planned during the follow-up period. The limited success in truly engaging and retaining families in the process of care is often the reason of failure to achieve desired health outcome.[37] Third, there was potential intervention contamination to the usual care group. Since families within the same community might be randomized to a different group, the

families in the usual care group might learn about the intervention from those assigned in the CTN group. The CTN trial was limited by the resource and funding to employ a large-scale multi-centre cluster design, which could protect against intervention contamination. Consequently, the contamination of control children reduced the estimate of intervention effect and could lead to a type II error on the results.[38] Fourth, our findings were limited by a relatively short period of observation after the initial launch of the intervention. For chronically ill children, the effect of integrated care on health outcomes would often accumulate over time. Thus, an interventional gain might be observed over a longer period of follow-up as the practitioners became more proficient and the team functioning got more collaborative. Fifth, our results were limited by missing data due to patients' loss to follow-up. We assumed that those data were MAR and employed multiple imputation in the analysis. However, the assumption of MAR is inherently untestable. Finally, the population targeted by the CTN may represent a subset of the defined CSHCN population accessing the services. Thus, our results may not be generalizable to a broader population of CSHCN, some of whom may not be accessing the services. The complex model of care and the heterogeneous population under study are the main challenges in assessing integrated health service interventions.[39]

Our findings should be interpreted in light of the previously published evidence on the CTN intervention. [22,23] For the subgroup of children who received complete CTN integrated care, their QOL was influenced by complex interactions between team integration and parenting factors; and overall, their QOL improved after 2 years.[22] In a cross-sectional analysis of the CTN children,[23] we found that the symptoms of psychiatric

distress experienced by the parents were associated with level of social support, family functioning, child behaviour, and parenting style. All those factors could be mediated through coordinated and family-centered care. Also, the costs of total health and social services were higher in those parents with more symptoms of psychiatric distress. From an economic perspective, continuous and comprehensive services can save costs because giving people what the services they need not only improves their QOL but also reduces the use of other duplicative or otherwise fruitless services.[40] A recent systematic review showed that coordinated and family-centered care was associated with improved outcomes for CSHCN.[41] A study by King et al.[42] showed that services for children with neurodevelopmental disorders were most beneficial when they were delivered in a family-centered fashion and addressed parent-identified issues. The current evidence is primarily from non-randomized studies. More RCTs are needed to assess the effect of integrative care that is continuous, comprehensive, coordinated, and family-centered care for CSHCN. The CTN intervention is the first of its kind in Canada. It aligns with the advocacy of comprehensive care for children's mental health,[43] family-centered rather than child-centered interventions,[44] and partnerships with community psychology in the system of care.[45] Considering that 13% to 18% of Canadian children had special health care needs, the CTN appears to be a milestone in the improvement of QOL of CSHCN through integrative care. At the same time, multiple community, family, parent, and child factors around the management of CSHCN make it challenging to evaluate this type of intervention. The on-going implementation and evaluation of the CTN is an important step forward for Canadian community-based childcare programs targeting this vulnerable

population. The methodological shortcomings in our study inform further trials to evaluate similar interventions.

Conclusion

In this trial, we assessed the effect of the CTN integrated care versus usual care on the psychosocial QOL of CSHCN in Simcoe County and York Region over 2 years. We did not find a significant improvement on the QOL for the children in the CTN group compared with those in the usual care group. The value of continuous, comprehensive, coordinated, and family-centered care for CSHCN is increasingly recognized. At the same time, more research is needed to demonstrate the magnitude of the effect of integrated care on those children and their parents. Given that non-compliance is a common issue in RCTs especially pragmatic trials, a better understanding on the performance of different analytical approaches is also crucial. Future simulation studies are needed to provide insights into this question.

Acknowledgments/Disclosures

Authors' contributions

CY conceived the sensitivity analysis, proposed statistical methods, performed all analyses, interpreted the results, and drafted and revised the manuscript. GB was commissioned by the CTN to design, carry out, analyze, and report the original trial. LT and JB contributed to the statistical design of the sensitivity analysis and revision of the manuscript. All authors have read and approved the final manuscript.

References

1. McPherson M, Arango P, Fox H, et al. A new definition of children with special health care needs. *Pediatrics* 1998, 102(1):137-139.

2. Bethell CD, Read D, Blumberg SJ, Newacheck PW. What is the prevalence of children with special health care needs? Toward an understanding of variations in findings and methods across three national surveys. *Maternal and child health journal* 2008, 12(1):1-14.

3. Gurney JG, McPheeters ML, Davis MM. Parental report of health conditions and health care use among children with and without autism: National Survey of Children's Health. *Archives of pediatrics & adolescent medicine* 2006, 160(8):825.

4. Ghandour RM, Kogan MD, Blumberg SJ, Perry DF. Prevalence and correlates of internalizing mental health symptoms among CSHCN. *Pediatrics* 2010, 125(2):e269-e277.

5. Patel V, Flisher AJ, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. *The Lancet* 2007, 369(9569):1302-1313.

6. Ingoldsby EM, Kohl GO, McMahon RJ, Lengua L. Conduct problems, depressive symptomatology and their co-occurring presentation in childhood as predictors of adjustment in early adolescence. *Journal of abnormal child psychology* 2006, 34(5):602-620.

7. Cossette L, Duclos É. A profile of disability in Canada 2001 (PALS). *Minister of Industry: Statistics Canada* 2001. http://www.statcan.gc.ca/pub/89-577-x/. Accessed May 20, 2013.

8. Simpson GA, Cohen RA, Bloom B, Blumberg SJ. The impact of children's emotional and behavioural difficulties on their lives and their use of mental health services. *Paediatric and perinatal epidemiology* 2009, 23(5):472-481.

9. Rosenzweig JM, Brennan EM, Huffstutter K, Bradley JR. Child care and employed parents of children with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders* 2008, 16(2):78-89.

10. Ball T. Learning from our past successes. *Children's Treatment Network Updates* 2013. http://www.ctn-simcoeyork.ca/resources/Learning%20From%20Our%20Past%20Successes.pdf. Accessed May 20, 2013.

11. Carnevale FA, Rehm RS, Kirk S, McKeever P. What we know (and do not know) about raising children with complex continuing care needs. *Journal of Child Health Care* 2008, 12(1):4-6.

12. The Well-being of Canada's young children report. *Government of Canada* 2011. http://www.dpe-agje-ecd-elcc.ca/eng/ecd/well-being/page11.shtml. Accessed May 20, 2013.

13. Hartley SL, Barker ET, Seltzer MM, et al. The relative risk and timing of divorce in families of children with an autism spectrum disorder. *Journal of family psychology: JFP: journal of the Division of Family Psychology of the American Psychological Association (Division 43)* 2010, 24(4):449.

14. Miller AR, Recsky MA, Armstrong RW. Responding to the needs of children with chronic health conditions in an era of health services reform. *Canadian Medical Association Journal* 2004, 171(11):1366-1367.

15. Ghandour RM, Perry DF, Kogan MD, Strickland BB. The medical home as a mediator of the relation between mental health symptoms and family burden among children with special health care needs. *Academic pediatrics* 2011, 11(2):161-169.

16. Committee MHIfCWSNPA. American Academy of Pediatrics. The medical home. *Pediatrics* 2002, 110(1 Pt 1):184-186.

17. Strickland BB, Singh GK, Kogan MD, Mann MY, van Dyck PC, Newacheck PW. Access to the medical home: new findings from the 2005-2006 National Survey of Children with Special Health Care Needs. *Pediatrics* 2009, 123(6):e996-e1004.

18. Starfield B, Shi L. The medical home, access to care, and insurance: a review of evidence. *Pediatrics* 2004, 113(Supplement 4):1493-1498.

19. Colpe LJ. Estimates of mental and emotional problems, functional impairments, and associated disability outcomes for the US child population in households. *Mental Health, United States* 2000:269-278.

20. Palfrey JS, Sofis LA, Davidson EJ, Liu J, Freeman L, Ganz ML. The Pediatric Alliance for Coordinated Care: evaluation of a medical home model. *Pediatrics* 2004, 113(Supplement 4):1507-1516.

21. Thurston S, Paul L, Browne G, et al. Early results: The 2-year comparative effect and expense of more and less integration of services that provide treatment and

rehabilitation for children with multiple disabilities: a randomized controlled trial. *Children's Treatment Network of Simcoe York, Final research report* March 2010.

22. Thurston S, Paul L, Ye C, et al. System integration and its influence on the quality of life of children with complex needs. *International Journal of Pediatrics* 2010, 2010.

23. Thurston S, Paul L, Loney P, Ye C, Wong M, Browne G. Associations and costs of parental symptoms of psychiatric distress in a multi-diagnosis group of children with special needs. *Journal of Intellectual Disability Research* 2011, 55(3):263-280.

24. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine* 2010, 8(1):18.

25. Simcoe York District Health Council Profile II. April 2002. http://www.ontla.on.ca/library/repository/mon/10000/250806.pdf. Accessed May 20, 2013.

26. Disability in Canada: A 2001 profile. *Office of Disabilities, Human Resources Development Canada* 2003. http://publications.gc.ca/collections/Collection/RH37-4-4-2001E.pdf. Accessed May 20, 2013.

27. Weiss ES, Anderson RM, Lasker RD. Making the most of collaboration: exploring the relationship between partnership synergy and partnership functioning. *Health Education & Behavior* 2002, 29(6):683-698.

28. Varni JW, Seid M, Rode CA. The PedsQL (TM): Measurement model for the pediatric quality of life inventory. *Medical care* 1999, 37(2):126-139.

29. Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Archives of general psychiatry* 1994, 51(1):8.

30. Strayhorn JM, Weidman CS. A parent practices scale and its relation to parent and child mental health. *Journal of the American Academy of Child & Adolescent Psychiatry* 1988, 27(5):613-618.

31. Cutrona CE, Russell DW. The provisions of social relationships and adaptation to stress. *Advances in personal relationships* 1987, 1(1):37-67.

32. Byles J, Byrne C, Boyle MH, Offord DR. Ontario Child Health Study: reliability and validity of the general functioning subscale of the McMaster Family Assessment Device. *Family process* 1988, 27(1):97-104.

33. Varni JW, Seid M, Kurtin PS. PedsQL (TM) 4.0: Reliability and validity of the pediatric quality of life Inventory (TM) Version 4.0 generic core scales in healthy and patient populations. *Medical care* 2001, 39(8):800-812.

34. Rubin DB. Multiple imputation for nonresponse in surveys, vol. 307: Wiley; 2009.

35. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000, 29(4):722-729.

36. Sheffler LC, Hanley C, Bagley A, Molitor F, James MA. Comparison of self-reports and parent proxy-reports of function and quality of life of children with below-the-elbow deficiency. *The Journal of Bone & Joint Surgery* 2009, 91(12):2852-2859.

37. Owens PL, Hoagwood K, Horwitz SM, et al. Barriers to children's mental health services. *Journal of the American Academy of Child & Adolescent Psychiatry* 2002, 41(6):731-738.

38. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ: British Medical Journal* 2001, 322(7282):355.

39. English M, Schellenberg J, Todd J. Assessing health system interventions: key points when considering the value of randomization. *Bulletin of the World Health Organization* 2011, 89(12):907-912.

40. Ye C, Browne G, Grdisa VS, Beyene J, Thabane L. Measuring the degree of integration for an integrated service network. *International Journal of Integrated Care* 2012, 12(5).

41. Kuhlthau KA, Bloom S, Van Cleave J, et al. Evidence for family-centered care for children with special health care needs: a systematic review. *Academic pediatrics* 2011, 11(2):136.

42. King G, King S, Rosenbaum P, Goffin R. Family-centered caregiving and well-being of parents of children with disabilities: linking process with outcome. *Journal of pediatric psychology* 1999, 24(1):41-53.

43. Tolan PH, Dodge KA. Children's mental health as a primary care and concern: A system for comprehensive support and service. *The American Psychologist* 2005, 60(6):601.

44. Briggs-Gowan MJ, Owens PL, Schwab-Stone ME, Leventhal JM, Leaf PJ, HoRWITZ

SM. Persistence of psychiatric disorders in pediatric settings. *Journal of the American

Academy of Child & Adolescent Psychiatry* 2003, 42(11):1360-1369.

45. Cook JR, Kilmer RP. Systems of care: New partnerships for community psychology.

American Journal of Community Psychology 2012, 49(3-4):393-403.

Appendix

Propensity score methods

Propensity score (PS) is the probability of receiving a treatment conditional on observed characteristics. Balancing on PS mimics randomization and produces comparable groups that are balanced on prognosis and observed confounding. We collected baseline variables that included: child's age, gender, admission diagnosis, parent's age, marital status, education, distress, parenting style, family income, social support, and family functioning. Using those variables, we calculated the PS of receiving the complete CTN integrated care by a logistic regression model. Four PS methods were used to balance between treated and untreated children in the AT and PP analyses: the matching, stratification, weighting, and covariate adjustment.

In the PS matching, we created matched pairs of treated and untreated children by matching them within 0.2 of the standard deviation of the logit of the PS. We used a 1:1 ratio to match the nearest children without replacement. This matching algorithm produced the least bias.[1] We then used the generalized estimating equation (GEE) model to analyze the matched data. An exchangeable correlation structure was used.

In the PS stratification, children were divided into five equal strata by the quintiles of their PS.[2] The outcome between treated and untreated children was compared directly in each stratum. The overall treatment effect is given by

$$\beta_{overall} = \frac{\sum_i^5 \beta_i}{5},$$

where β_i is the estimated treatment effect in stratum i. The variance of $\beta_{overall}$ is calculated by

$$\text{Var}(\beta_{\text{overall}}) = \frac{1}{\Sigma_i^5 w_i} \text{,}$$

where w_i is the inverse of the estimated variance of β_i for stratum i.

In the PS-weighted method, we balanced the children by the inverse probability of receiving the CTN integrated care. The weight (w_i) for each child can be calculated by

$$w_i = \frac{X_i}{e_i} + \frac{(1-X_i)}{1-e_i},$$

where X_i is the treatment indicator; and e_i is the estimated PS for child i. Confounding between the observed variables and the treatment was eliminated in the weighted sample. We then directly compared the outcome between treated and untreated children by a weighted linear regression model.

In the covariate adjustment method, we adjusted the PS as the sole covariate in the regression model for estimating the treatment effect.

Calculating the standard error for instrumental variable estimate

Suppose we have two random variables X and Y. A Taylor series expansion of $f(x, y)$ about the values (x_0, y_0) is given by

$$f(x, y) = f(x_0, y_0) + (x - x_0) \frac{\partial f(x, y)}{\partial x}\bigg|_{(x_0, y_0)} + (y - y_0) \frac{\partial f(x, y)}{\partial y}\bigg|_{(x_0, y_0)} + O(\cdot)$$

$$\approx f(x_0, y_0) + (x - x_0) \frac{\partial f(x, y)}{\partial x}\bigg|_{(x_0, y_0)} + (y - y_0) \frac{\partial f(x, y)}{\partial y}\bigg|_{(x_0, y_0)}$$

where $O(\cdot)$ is the higher order terms which are omitted in this approximation.

The instrumental variable (IV) estimator for the exposure X and the outcome Y is given by

$$\beta_{IV} = \frac{\beta_{(Z \to Y)}}{\beta_{(Z \to X)}}$$

where $\beta_{(Z \to Y)}$ and $\beta_{(Z \to X)}$ are the parameters of association between Z and Y and between Z and X respectively. By the Taylor series expansion on β_{IV}, we have

$$\beta_{IV} \approx \frac{b_{(Z \to Y)}}{b_{(Z \to X)}} + \left(\beta_{(Z \to Y)} - b_{(Z \to Y)}\right)\frac{1}{b_{(Z \to X)}} - \left(\beta_{(Z \to X)} - b_{(Z \to X)}\right)\frac{b_{(Z \to Y)}}{b_{(Z \to X)}^2}$$

where $b_{(Z \to Y)}$ and $b_{(Z \to X)}$ are two values at which β_{IV} is differentiable. Then the variance of the IV estimator can be approximated as

$$\text{Var}(\beta_{IV}) \approx \frac{1}{b_{(Z \to X)}^2}\text{Var}\left(\beta_{(Z \to Y)}\right) + \frac{b_{(Z \to Y)}^2}{b_{(Z \to X)}^4}\text{Var}\left(\beta_{(Z \to X)}\right) - \frac{2b_{(Z \to Y)}}{b_{(Z \to X)}^3}\text{Cov}\left(\beta_{(Z \to X)}, \beta_{(Z \to Y)}\right).$$

Under the assumption that $\beta_{(Z \to X)}$ is independent of $\beta_{(Z \to Y)}$, the variance of the IV estimator is then

$$\text{Var}(\beta_{IV}) \approx \frac{1}{b_{(Z \to X)}^2}\text{Var}\left(\beta_{(Z \to Y)}\right) + \frac{b_{(Z \to Y)}^2}{b_{(Z \to X)}^4}\text{Var}\left(\beta_{(Z \to X)}\right).$$

We substitute $b_{(Z \to Y)}$ and $b_{(Z \to X)}$ by the estimates of $\beta_{(Z \to Y)}$ and $\beta_{(Z \to X)}$ respectively, and approximate the $\sqrt{\text{Var}\left(\beta_{(Z \to Y)}\right)}$ and $\sqrt{\text{Var}\left(\beta_{(Z \to X)}\right)}$ by the associated standard error (SE).

Thus, we can obtain an approximate variance of the IV estimate. In our analysis, we used the least squares estimate of $\beta_{(Z \to Y)}$ and associated SE obtained from the linear regression model. For the exposure X (a binary indicator of whether or not a patient received complete CTN integrated care), the association between X and the IV can be calculated by

$$\beta_{(Z \to X)} = P(X = 1|Z = 1) - P(X = 1|Z = 0)$$

where $P(X = 1|Z = 1)$ represents the proportion of treated patients in the CTN group;

and $P(X = 1|Z = 0)$ is always zero because the children in the usual care group are

deemed to be untreated. The variance of $\beta_{(Z \rightarrow X)}$ is

$$\text{Var}(\beta_{(Z \rightarrow X)}) = \frac{P(X = 1|Z = 1)(1 - P(X = 1|Z = 1))}{n}$$

where n is the number of children in the CTN group.

References

1. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics* 2011;10:150-161.

2. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 1984;79(387):516-524.

Figure 1: Components of the Children's Treatment Network versus usual care

Timeline	CTN integrated care	Usual care
Start up	a b c d e	b
Recruitment, baseline measures, randomization	f	
During 24 months	g h i	i
At 12 and 24 months	Measurement of outcomes	

a	Government funded the CTN: board established, executive and associate Directors hired, integrated service agreements accomplished, meetings with network partners to monitor ongoing network development and address challenges
b	Government funded to enhance rehab and mental health services throughout Simcoe County and York Region.
c	Single point of access: the CTN established the access form and hired workers called "system navigators", who referred information to the child and family coordinator.
d	The CTN established 10 local teams made up of interdisciplinary clinicians from participating partners: 1) training in creating a single plan of care; 2) training in integrated service team functioning; and 3) training team coordinators
e	The CTN set up and trained the use of the E-Record.
f	REB approved the RCT. Investigators pilot-tested recruitment, enrolled children, conducted baseline measures, and randomized children into either group.
g	The CTN hired and trained team facilitators. A child and family team for each case was formed from members of local teams. The CTN assessed the integration of partners and the functioning of child and family teams.
h	• **Single point of access**: one number to call to access any service for child or family. • **System navigator**: did intake; set up e-record; assessed problems; and identified regional team coordinators. • **Team coordinator**: assembled team members and skills to match child and family needs; conducted team meetings; and arranged service sequence and timing. • **Child and family team**: agreed on goals and single plan of care for child and requested services. • **Electronic record**: single point of access to obtain and record information; common record available to all team members to record notes and progress
i	Usual wait-list as a function of prioritization

☐ = fixed objects   ◯ = flexible activities

Figure 2: Summary of different analytical approaches
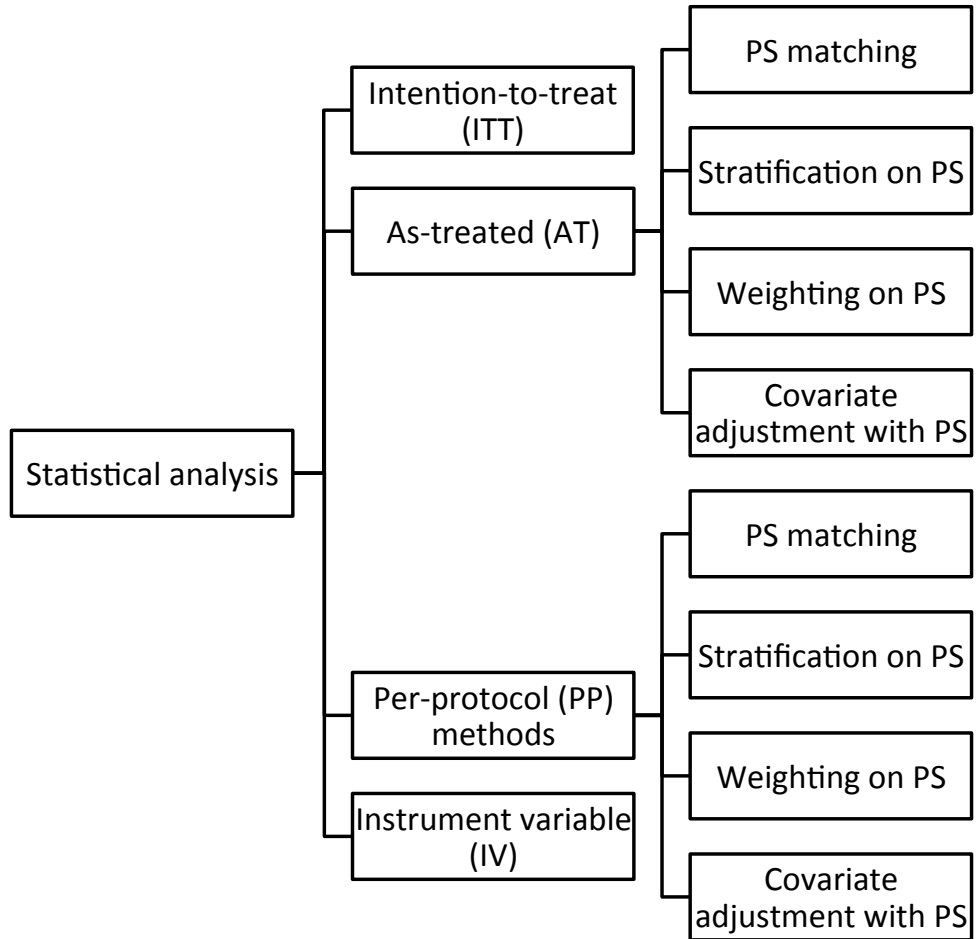
Figure 3: Flow of the children in the Children's Treatment Network Trial

Assessed for eligibility (n= 2319)

Excluded (n=1874):
- ◆ No response; declined to participate
- ◆ Did not complete baseline interview

Randomization (n=445)

Allocated to intervention (n=229)
- ◆ Treated children (n=111)
- ◆ Untreated children (n=118):
  1. Ongoing integration (n=7)
  2. Coordinated but did not receive services (n=53)
  3. Declined CTN services (n=58)

Allocated to control (n=216)
- ◆ Received usual care (n=216)

Lost to follow-up (n=64):
- 24 from the treated
- 40 from the untreated

Lost to follow-up (n=57)

Analysis:
- ◆ Intention-to-treat analysis: (n=229)
- ◆ As-treated analysis: (n=229: 111 treated ,118 untreated)
- ◆ Per-protocol analysis: (n=111)
- ◆ Instrumental variable analysis: (n=229)

Analysis:
- ◆ Intention-to-treat analysis: (n=216)
- ◆ As-treated analysis: (n=216; all untreated)
- ◆ Per-protocol analysis: (n=216)
- ◆ Instrumental variable analysis: (n=216)

Figure 4: Comparing the estimates of treatment effect from different analyses
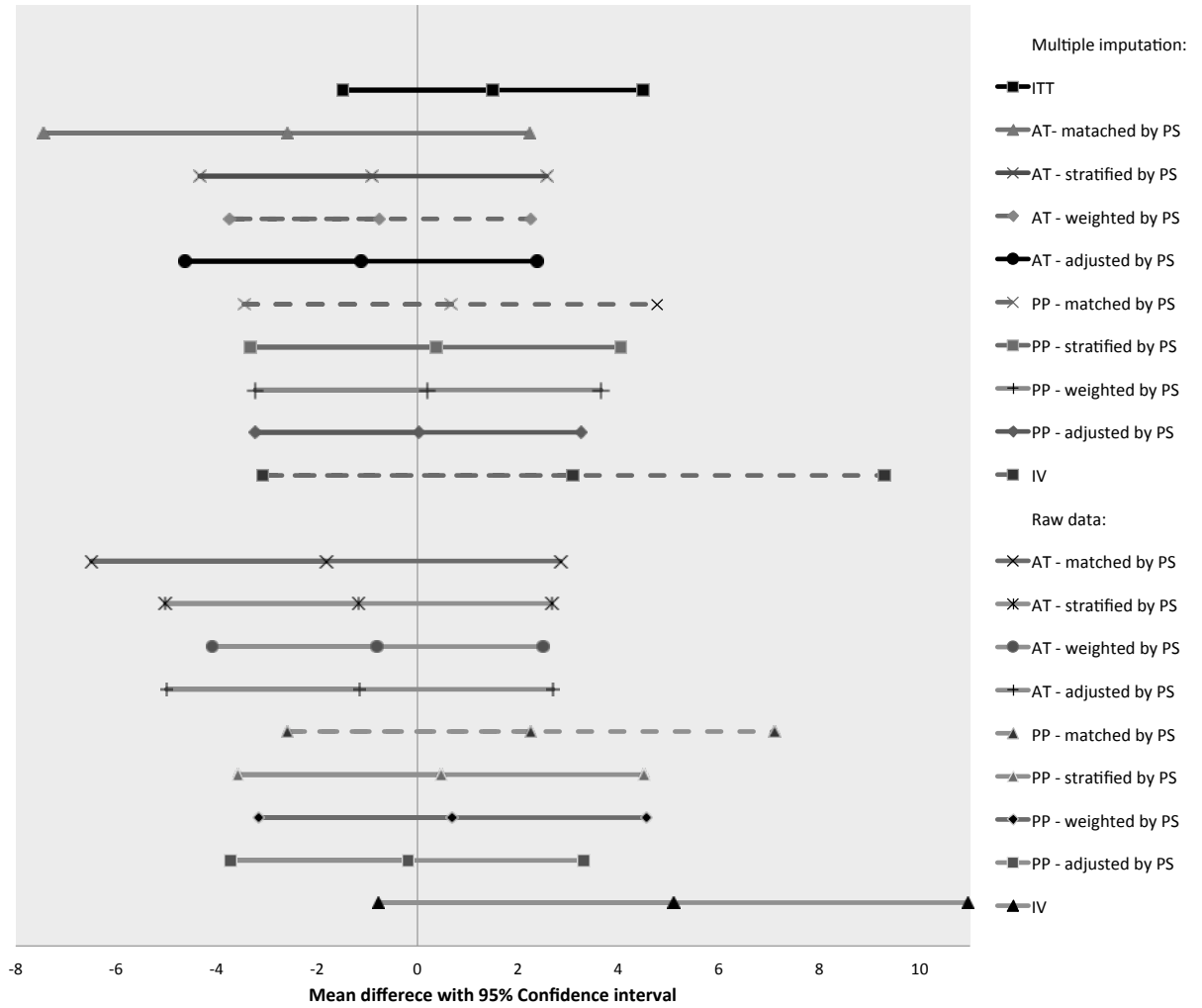
ITT= intention-to-treat; AT= as-treated; PP= per-protocol; IV= instrumental variable; PS= propensity score; Mean difference = the difference on the change of psychosocial score in 2 years between groups.

Table 1: Baseline characteristics

Baseline variable	Intervention (n=229)	Control (n=216)	p value
Child's age: mean (SD)	7.8 (4.3)	8.1 (4.6)	0.60
Child's gender: frequency (%)			
Male	148 (64.6%)	149 (69.0%)	0.33
Female	81 (35.4%)	67 (31.0%)	
Child's admission diagnostic: frequency (%)			
Mental and developmental disorders	106 (46.3%)	95 (44.0%)	0.81
Diseases of the nervous system	53 (23.1%)	51 (23.6%)	
Congenital abnormalities	39 (17.0%)	34 (15.7%)	
Other disease	31 (13.5%)	36 (16.7%)	
Child's psychosocial score: mean (SD)	59.0 (18.6)	59.2 (18.6)	0.85
Parent's age: mean (SD)	40.5 (7.6)	40.4 (7.7)	0.95
Marital status: frequency (%)			
Married including common-law	190 (83.0%)	187 (86.6%)	0.29
Other	39 (17.0%)	29 (13.4%)	
Parent's education: frequency (%)			
Secondary	83 (36.4%)	78 (36.1%)	0.95
Post-secondary	145 (63.6%)	138 (63.9%)	
Family annual income: frequency (%)			
Less than $30,000	32 (14.0%)	32 (14.9%)	0.96
$30,000 to $90,000	121 (53.1%)	114 (53.0%)	
Above $90,000	75 (32.9%)	69 (32.1%)	
Parent's Kessler distress score: mean (SD)	19.5 (5.8)	20.4 (7.2)	0.16
Positive parenting score: mean (SD)	15.2 (3.1)	15.1 (3.0)	0.82
Social support score: mean (SD)	17.6 (4.7)	17.5 (4.3)	0.74
Family functioning score: mean (SD)	9.1 (6.3)	9.4 (5.9)	0.60

The continuous and categorical variables are expressed in mean (standard deviation) and frequency (percentage), respectively.

SD = standard deviation. The p value was calculated based on a t-test for continuous variables and a chi-square test for categorical variables.

Table 2: Comparing the treated with untreated children within the intervention group

Baseline variable	Treated children (n=111)	Untreated children (n=118)	p value
Child's age: mean (SD)	7.3 (4.3)	8.3 (4.3)	0.09
Child's gender: frequency (%)			
Male	70 (63.1%)	78 (66.1%)	0.63
Female	41 (36.9%)	40 (33.9%)	
Child's admission diagnosis: frequency (%)			
Mental and developmental disorders	46 (41.4%)	60 (50.8%)	0.13
Diseases of the nervous system	32 (28.8%)	21 (17.8%)	
Congenital abnormalities	21 (18.9%)	18 (15.3%)	
Other disease	12 (10.8%)	19 (16.1%)	
Child's psychosocial score: mean (SD)	58.1 (19.0)	59.9 (18.4)	0.49
Parent's age: mean (SD)	40.4 (6.9)	40.6 (8.3)	0.81
Marital status: frequency (%)			
Married including common-law	92 (82.9%)	98 (83.1%)	0.97
Other	19 (17.1%)	20 (16.9%)	
Parent's education: frequency (%)			
Secondary	40 (36.0%)	43 (36.8%)	0.91
Post-secondary	71 (64.0%)	74 (63.2%)	
Family annual income: frequency (%)			
Less than $30,000	19 (17.3%)	13 (11.0%)	0.34
$30,000 to $90,000	58 (52.7%)	63 (53.4%)	
Above $90,000	33 (30.0%)	42 (35.6%)	
Parental Kessler distress score: mean (SD)	20.2 (5.9)	18.8 (5.6)	0.09
Parenting style score: mean (SD)	15.7 (3.2)	14.8 (3.0)	0.02
Social support score: mean (SD)	17.7 (4.6)	17.6 (4.9)	0.86
Family functioning score: mean (SD)	9.1 (6.4)	9.1 (6.3)	0.97

The continuous and categorical variables were expressed in mean (standard deviation) and frequency (percentage), respectively.

SD = standard deviation. The p value was calculated based on a t-test for continuous variables and a chi-square test for categorical variables.

Table 3: Summary of the estimates of treatment effect

	Multiple imputation			Raw data		
	MD	95% CI	p value	MD	95% CI	p value
ITT analysis[a]	1.50	-1.49 4.50	0.32	-[a]	- -	-
AT analysis						
Matched by PS	-2.60	-7.45 2.24	0.29	-1.82	-6.49 2.86	0.45
Stratified by quintiles of PS	-0.89	-4.34 2.57	0.61	-1.17	-5.01 2.68	0.55
Weighted by PS	-0.75	-3.75 2.25	0.62	-0.80	-4.10 2.50	0.63
Adjusted by PS	-1.12	-4.62 2.39	0.53	-1.15	-4.99 2.70	0.56
PP analysis						
Matched by PS	0.67	-3.44 4.78	0.75	2.26	-2.60 7.12	0.36
Stratified by quintiles of PS	0.37	-3.33 4.06	0.85	0.47	-3.58 4.52	0.82
Weighted by PS	0.21	-3.24 3.66	0.91	0.70	-3.16 4.56	0.72
Adjusted by PS	0.02	-3.23 3.27	0.99	-0.20	-3.71 3.31	0.91
IV analysis	3.10	-3.08 9.29	0.33	5.10	-0.78 10.97	0.09

[a]The ITT analysis required complete patients and was only performed in multiple imputation.
MD = mean difference, the difference on the change of psychosocial score in 2 years between groups; CI = confidence interval.
ITT= intention-to-treat; AT= as-treated; PP= per-protocol; IV= instrumental variable; PS= propensity score.

CHAPTER 4

Estimating treatment effects in randomized controlled trials with non-compliance: a simulation study

Chenglin Ye[1,2], Joseph Beyene[1], Gina Browne[1,3], Lehana Thabane[1,2*]

Affiliations:

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton ON, L8S 4L8 Canada

[2]Biostatistics Unit, St. Joseph's Healthcare Hamilton

[3]School of Nursing, McMaster University, Hamilton ON, L8S 4L8 Canada

[*]Corresponding author:

Lehana Thabane

Email: thabanl@mcmaster.ca

Phone: +1/905-522-1155 ext.33720

Fax: +1/905-308-7212

Abstract

Objective: Randomized controlled trials (RCTs) have been the gold standard to assess

new health interventions. Patients are randomly assigned to receive an intervention or

control. The effect of an intervention can be estimated by comparing outcomes between

groups, whose prognostic factors are balanced by randomization. However, patients' non-

compliance with the assigned treatment undermines randomization and will potentially

lead to a biased estimate of treatment effect. Through simulation, we aim to compare

some common approaches in analyzing non-compliant data under different non-compliant

scenarios.

Settings: Based on a real study, we simulated hypothetical trials by varying three non-

compliant factors: the type, randomness, and degree of non-compliance. We employed

the intention-to-treat (ITT), as-treated (AT), per-protocol (PP), instrumental variable (IV),

and complier average casual effect (CACE) analyses to estimate large (50% improvement

over the control), moderate (25%), and null (0%) treatment effects. The results were

compared by the bias of estimates, mean square error, and 95% coverage of the true

value.

Results: For a large or moderate treatment effect, the ITT estimate was considerably

biased in all scenarios. The AT, PP, IV, and CACE estimates were unbiased when non-

compliant behaviours were random. The IV estimate was unbiased when non-compliant

behaviours were symmetrically associated with patients' inherent conditions. The PP

estimate was unbiased when non-compliers had no alternative treatments besides the

control.

Conclusion: The standard ITT analysis under non-compliance is biased when

interventions have moderate or large effect. Alternative analyses can provide unbiased or

less biased estimates. Based on the results, we make some suggestions on choosing

optimal approaches of analyzing non-compliant data for specific non-compliant scenarios.

Keywords: randomized controlled trial, non-compliance, intention-to-treat, as-treated,

per-protocol, instrumental variable, complier average causal effect

Background

Randomized controlled trials (RCTs) have been the gold standard to assess new health interventions. Patients are randomly assigned to receive an intervention or control (usually the standard treatment). Randomization on average balances different treatment groups on their prognostic factors so observed difference can be attributed to the effect of different treatments. A common challenge in analyzing RCTs is to deal with patients' non-compliance. Non-compliance is the case when a patient does not fully comply with the treatment that he or she was assigned to. Non-compliance compromises the 'fair' comparison between treatment groups, which was protected by randomization, and potentially leads to a biased estimate of treatment effect. While different analyses have been used to deal with non-compliance, the bias of treatment estimate is rarely studied. Result interpretations also vary depending on the nature of non-compliance and the objective of a study. Some RCTs, known as pragmatic trials [1-3], are primarily designed to guide clinical practice. Their goal is often to assess if a new intervention works in routine practice. For other non-pragmatic trials, the goal is normally to study the biological efficacy of an intervention. Despite the objective of a RCT, an analysis that provides an unbiased or less biased estimate of treatment effect is always desirable. In this study, we compare some common analyses to deal with different non-compliant scenarios. The results will provide useful knowledge in choosing optimal methods to analyze non-compliant data.

This study was motivated by a RCT that compared the integrated care organized through the Children's Treatment Network (CTN) with the usual care directed by parents

for managing children with special health care needs [4]. The CTN coordinated

community resources to deliver comprehensive health services for the target children and

their families. The hypothesis was that the target children and their parents' health

outcomes would improve by getting integrated, proactive, and necessary services for their

families. While the use of RCTs in assessing CTN-like interventions has been promising,

non-compliant rates are usually high in those RCTs. That is largely due to the complexity

of implementing those interventions in real-life settings.

An 'as randomized', or intention-to-treat (ITT), analysis is commonly used in RCTs

as the gold standard [5]. Yet, excessive reliance on the ITT analysis can be problematic

for non-compliant scenarios [6]. It has been suggested [7,8] that all RCTs with non-

compliance need alternative analyses. As-treated (AT) and per-protocol (PP) analyses are

two common alternatives. Different from the ITT analysis, both AT and PP analyses do

not preserve original randomization and are prone to prognostic imbalance. The

instrumental variable (IV) analysis, which uses randomization allocation to adjust for

confounding, is another approach to deal with non-compliance. The complier average

causal effect (CACE) analysis has also been proposed [9] to estimate treatment effects

among compliers. In this study, we compared all those five approaches, namely, the ITT,

AT, PP, IV, and CACE in analyzing different non-compliant scenarios.

Our objective was to compare the analyses by the bias of estimated treatment effect,

the mean square error (MSE), and the 95% coverage of the true value. The data were

generated through hypothetical RCTs with different non-compliant scenarios. A similar

simulation study was conducted previously to compare the ITT, AT, PP, and IV [10]. The

authors showed that the ITT and IV analyses carried desirable properties but were biased in certain cases. Our proposed simulation included the CACE method and compared additional non-compliant scenarios. For example, previous study did not consider the case of no crossover between treatment groups. That was the case when an intervention was only accessible to patients who were offered it. Our study would provide further insights into the analyses of RCTs with non-compliance.

Methods

Simulation framework

In the CTN trial, over 50% of the children randomized in the CTN group did not fully comply with the intervention for various reasons. Both primary and sensitivity analyses showed that the effect of the CTN was not significant but the estimates varied in direction, magnitude, and precision [4]. That observation prompted us to further investigate the impact of non-compliance on estimated treatment effects.

Based on the CTN setting, we simulated hypothetical RCTs where patients were randomly assigned to an intervention or usual care by a 1:1 allocation ratio. The parameters for generating hypothetical patients were estimated from the CTN trial. We generated different non-compliant scenarios by varying three factors: 1) the type of non-compliers; 2) the randomness of non-compliance; and 3) the degree of non-compliance. Our simulation framework is shown in Figure 1. The design, conduct, and reporting of this study follow the guideline of designing and reporting simulation studies [11].

Type of non-compliers

We considered two types of non-compliers which were defined as: never-takers and always-takers [12]. Never-takers are the patients who will always reject the new intervention if they were offered it. Always-takers will always receive the new intervention even they were not offered it. Two scenarios were considered. In scenario one, we assumed that non-compliers were either never-takers or always-takers. That mimicked the situation when patients were able to get the intervention elsewhere even they were not assigned to it. In scenario two, we assumed that non-compliers were only never-takers. That mimicked the situation when a new intervention was only accessible to patients who were assigned to it. Additionally, we assumed that non-compliers who did not get the intervention would receive the usual care instead.

Randomness of non-compliance

Non-compliant behaviours could be purely random or dependent on patients' characteristics. For the dependent case, we assumed that non-compliant behaviours were dependent on patients' conditions. In particular, we considered six non-compliant scenarios that were studied by McNamee [13], including:

A. patients with good conditions would always get the intervention while patients with poor conditions would always reject it;

B. patients with good conditions would always get the intervention;

C. patients with poor conditions would always reject the intervention;

D. patients with good conditions would always reject the intervention while patients with

poor conditions would always get it;

E. patients with good conditions would always reject the intervention;

F. patients with poor conditions would always get the intervention.

We assumed that patients' conditions were positively associated with their outcomes

when they received usual care. Thus, patients' conditions were defined to be good when

they had an outcome score at least 0.5 standard deviations above the group mean under

usual care (considering that a high score was a better outcome). Similarly, patients'

conditions were defined to be poor when they had have an outcome score at least 0.5

standard deviations below the group mean under usual care. When there were only never-

takers, only assumptions C and E were considered.

Degree of non-compliance

The degree of non-compliance referred to the proportion of intervention that non-

compliers did not receive. The simplest case was all-or-none where compliers received

100% of the intervention and non-compliers received none of it. For interventions like the

CTN integrated care, patients were likely to receive some components of the intervention

even they did not fully comply with it. In addition, intervention fidelity might also

contribute to partial non-compliance. A systematic review showed that many

interventions of integrated care failed to achieve a full integration of services as planned

[14]. Both all-or-none and partial non-compliance were considered in our simulation.

For all-or-none case, we considered only two non-compliant levels: none ($d = 0$) or all components of the intervention ($d = 1$). A study reported that non-compliant rate could be as high as 30-40% for a treated population [15]. Thus, we randomly selected 30% patients to receive the treatment opposite to what they were assigned for. For example, if selected patients were assigned to the intervention, they would receive none of it ($d = 0$). For partial non-compliance, we considered four levels: none ($d = 0$), one third ($d = \frac{1}{3}$), two thirds ($d = \frac{2}{3}$), or all ($d = 1$) components of the intervention. Those four levels were proposed in a previous simulation study conducted by Bang & David [10].

Simulation procedures

To generate the data, we employed a modified model from a previous simulation study [10]. Let Y_1 and Y_0 be a pair of counterfactual outcomes for a patient if he or she were in the intervention and the usual care groups, respectively. We then generated the outcome through a Normal distribution

$$Y_k \sim Normal(\mu_k, 10^2) \text{ and } k = 0,1,$$

where we chose $\mu_0 = 59$ for the usual care group; and $\mu_1 = 89$, 74 and 59 for the intervention group. The mean score of 59 for the usual care group and the standard deviation of 10 were estimated from the CTN trial. By adopting a marginal view, we defined the treatment effect δ between groups as

$$\delta = \mu_1 - \mu_0,$$

where μ_1 and μ_0 were the mean values of Y_1 and Y_0, respectively. The mean scores of the intervention were chosen to have a large (50% improvement over the usual care group), moderate (25%), or null (0%) effect. A group indicator Z (1 = intervention and 0 = usual care) was generated for each patient from a Bernoulli distribution with equal probability of 0.5 being assigned in either group. Assuming that the outcome score was proportional to the actual level of treatment received, the observed outcome for patient i was calculated by

$$y_i = d_i Y_1 + (1 - d_i) Y_0,$$

where d_i was the proportion of intervention that patient i received.

In the CTN trial, 450 patients were needed to detect a minimum clinically important difference (MCID) of 15 with 80% statistical power and 5% alpha. Using the same MCID, we chose to simulate 500 subjects in each hypothetical trial. We estimated the standard error of the treatment effect to be 1.53 from the CTN trial. Based on that estimate, at least 816 simulations were needed to produce an effect estimate within 1% accuracy of the MCID by the standard formula [11]. To have sufficient power, we chose to generate 1000 simulations per scenario. The steps of simulation are shown in Figure 2.

Statistical analysis

This section describes different analyses that we studied. The estimated treatment effect was expressed as the difference of the mean score between groups. The methods were compared by the bias, MSE, and 95% coverage of the true value by following the guideline of reporting simulation studies [11].

Intention-to-treat (ITT)

In the ITT method, the outcome was compared between groups by randomization regardless of whether they were compliant or not. The effect estimate (δ) was:

$$\delta_{ITT} = \bar{y}_{Z=1} - \bar{y}_{Z=0},$$

where $\bar{y}_{Z=1}$ and $\bar{y}_{Z=0}$ were the mean outcome scores of the intervention and usual care groups, respectively.

As-treated (AT)

The AT approach compared the patients by the treatment they actually received. The effect estimate was:

$$\delta_{AT} = \bar{y}_{d=1} - \bar{y}_{d=0},$$

where d indicated the proportion of intervention that a patient had received. For all-or-none case, d was either 1 or 0 and the AT estimate was calculated by directly subtracting two means. For partial non-compliance case, d took a value of $0, \frac{1}{3}, \frac{2}{3}$, or 1 and δ_{AT} was obtained by regressing the variable d on the outcome in a linear regression model [10].

Per-protocol (PP)

The PP method excluded patients who did not fully comply with their assigned treatment. The effect estimate was:

$$\delta_{PP} = \bar{y}_{Z=d=1} - \bar{y}_{Z=d=0}.$$

Instrumental variable (IV)

The IV approach employed the randomization indicator Z as an instrumental variable to adjust for the proportion of non-compliant patients. The theory and

assumptions of IV analysis were thoroughly discussed in the literature [16,17]. We used

the standard IV estimate:

$$\hat{\delta}_{IV} = \frac{\bar{y}_{Z=1} - \bar{y}_{Z=0}}{\bar{d}_{Z=1} - \bar{d}_{Z=0}}.$$

The Fieller's theorem was used to calculate the standard error of the estimate [10].

Complier average causal effect (CACE)

The CACE method estimated the treatment effect among compliers and the estimate

was

$$\delta_{CACE} = E[\bar{y}_{Z=1} - \bar{y}_{Z=0}|compliers].$$

A challenge in CACE estimation was to categorize compliance status for patients whose

non-compliance could not be observed. For example, some non-compliers in the usual

care group might appear to be compliant because they were not accessible to other

interventions besides the usual care. To handle this problem, those unknown compliance

cases were treated as missing and estimated by using the expectation-maximization (EM)

algorithm [18,19]. The CACE analysis was performed in Mplus (version 7) [Mac OS X

10.6.8] Los Angeles, CA: Muthén & Muthén. The rest analyses and all simulations were

performed in R version 2.15.2.

Cut-off points for non-compliance

In practice, investigators often dichotomize patients to be either a complier or non-

complier. A cut-off of 80% is commonly used [20,21]. Patients are considered to be

compliers if they have complied with at least 80% of the intervention. A cut-off of 100%

has also been used such that patients are only considered to be compliers if they have

received all of the intervention. Generally, compliers are assumed to benefit from the

complete effect of an intervention. We conducted a sensitivity analysis to investigate the impact of those two cut-offs on estimating treatment effects. A new compliance indicator for patient i was defined as: $t_i = I(d_i \geq 0.8)$ for a cut-off of 80% and $t_i = I(d_i = 1)$ for a cut-off of 100%. The indicator function I returned 1 if the condition was satisfied and 0 otherwise. We then performed the same analyses by replacing d_i with t_i for patient i.

Results

The estimates by different analyses under the simulated scenarios, and their bias, MSE, and 95% coverage are reported in Tables 1 to 6. For a large or moderate intervention effect (a mean difference of 30 and 15 representing 50% and 25% improvement over the usual care, respectively), the ITT estimate was considerably biased (Figures 3 and 4). The other estimates were unbiased when non-compliant behaviours were random. When non-compliant behaviours were not random, the PP estimate was mostly unbiased if non-compliers were all never-takers. When there were both always- and never-takers, the PP and CACE estimates were the least biased if never-taking behaviours were dependent on patients' conditions (dependent non-compliance scenarios C and E). The IV estimate was the least biased if non-compliant behaviours were symmetrically dependent on patients' conditions (dependent non-compliance scenarios A and D). When there were only never-takers, the PP estimates were mostly unbiased.

For a null treatment effect (no improvement to the usual care), all estimates were unbiased when non-compliant behaviours were random (Figure 5). The ITT estimate was also unbiased if non-compliant behaviours were symmetrically dependent on patients'

conditions. In other cases, all estimates were biased to some extent but the ITT was the

least biased. We noticed that the IV estimate generally had a larger MSE than the other

estimates. That was due to the fact that the IV estimate was more sensitive to non-

compliant rate and would approach infinity as non-compliant rate was getting to be equal

between treatment groups. The MSE of all estimates are shown in Figures 6 to 8 and

follows the same pattern as the bias.

When both never- and always-takers were considered, both ITT and IV approaches

showed a better 95% coverage of the true value than the other approaches. For the large

treatment effect, the 95% coverage was 0% for the ITT approach in all scenarios. The IV

approach generally showed a better 95% coverage than the other methods. When only

never-takers were allowed, the PP estimate showed the best 95% coverage among all

methods.

In the sensitivity analysis, we compared the impact of using a cut-off of 80% or

100% on dichotomizing patients' non-compliance status according to their degree of

adherence. The results showed that dichotomizing patients by a cut-off of 80% produced

less biased results than dichotomizing them by a cut-off of 100%. For a null treatment

effect, the treatment estimates obtained by applying a cut-off of 80% were even less

biased than analyzing patients by their actual degree of compliance.

Discussion

Through simulation, we compared different methods of analyzing non-compliant

RCT data. Our results showed that the ITT approach was the most optimal when

estimating a null effect since it provided an unbiased or the least biased estimate in different scenarios. This result was consistent with the general opinion that the ITT estimate was conservative towards the null. However, for the case of a large or moderate treatment effect, the ITT approach was much more biased than the other approaches. When patients' non-compliant behaviours were purely random, the AT, PP, IV, and CACE approaches all provided unbiased estimates. For other non-compliant behaviours that we considered, the choice of optimal method varied. Figure 9 summarizes the choices of optimal methods under different scenarios to produce an unbiased or less biased estimate.

Although the ITT method is the most commonly reported analysis, other analyses of non-compliant data may provide a better estimate. Thus, understanding which analyses are unbiased or less biased is important in choosing the analyses and interpreting the results. Our results are limited by a number of factors. Firstly, we did not consider other prognostic factors in the simulation. Adjusting for prognostic factors can affect the estimation of treatment effect. Secondly, we assumed that the effect of intervention was proportional to the actual level of treatment received in our simulation. That association could be more complicated in real life. Thirdly, we did not consider patients' withdrawal from the study and assumed that non-compliers' outcomes were still collectable. Finally, we only simulated a subset of general non-compliant scenarios. Thus, our findings may not be generalizable to other scenarios.

Despite the limitations, our study has several strengths. The simulation framework was built on three key factors of non-compliance: the type of non-compliers, the

randomness of non-compliance, and the degree of non-compliance. Those three factors were hardly considered simultaneously in previous studies. We generated a total of 30 non-compliant scenarios by varying the three factors and the effect size of treatment. The findings help investigators choose the optimal approaches when analyzing similar non-compliant problems. Our results also confirm some previous conclusions from dealing with non-compliance. For example, the ITT analysis was unbiased if the treatment effect was zero [22]. All estimates were unbiased if non-compliance was independent of patients' outcomes and the IV estimate was also unbiased when non-compliance was symmetrically dependent on patients' outcomes [10]. In addition, we found that the PP estimate was unbiased when there were only never-takers. Dealing with issues of non-compliance is challenging. The real impact of non-compliance in statistical analysis is difficult to assess. We have compared the performance of common analyses under specific non-compliant scenarios. The results highlight the value of employing multiple approaches to analyze non-compliant data. Our work also contributes to the quality assessment of research evidence generated from RCTs subject to non-compliance and provides basis for a more complex evaluation.

Conclusion

Our simulation shows that the ITT analysis under non-compliance is considerably biased when an intervention has a large effect over the control. Alternative analyses can provide unbiased or less biased estimates. For RCTs subject to substantial non-compliance, we make some suggestions for the choice of analyses under specific

scenarios to minimize the bias of estimated treatment effects. Our study informs the

design of further investigations on the issue of non-compliance in RCTs.

Authors' contributions

CY conceived the study, designed and performed the simulations, conducted the statistical analyses, interpreted the results, and drafted the manuscript; LT advised on the design of the study and revised the manuscript; JB and GB contributed to the interpretation of the results and revision of the manuscript; all authors have read and approved the final manuscript.

Acknowledgements

References

1. McMahon AD: Study control, violators, inclusion criteria and defining explanatory and pragmatic trials. *Statistics in medicine* 2002, 21(10):1365-1376.

2. Schwartz D, Lellouch J: Explanatory and pragmatic attitudes in therapeutical trials. *Journal of clinical epidemiology* 2009, 62(5):499-505.

3. Tunis SR, Stryer DB, Clancy CM: Practical clinical trials. *JAMA: the journal of the American Medical Association* 2003, 290(12):1624-1632.

4. Ye C, Browne G, Beyene J, Thabane L: A sensitivity analysis of the Children's Treatment Network trial: a randomized controlled trial of integrated services versus usual care for children with special health care needs. *Clinical Epidemiology* 2013, 5:373-385.

5. Hubbard W: ICH: Guidance on Statistical Principles for Clinical Trials (FDA Docket No. 97D-0174). *Federal Register* 1998, 63:49583-49598.

6. Sheiner LB, Rubin DB: Intention-to-treat analysis and the goals of clinical trials&ast. *Clinical Pharmacology & Therapeutics* 1995, 57(1):6-15.

7. Nich C, Carroll KM: 'Intention-to-treat' meets 'missing data': implications of alternate strategies for analyzing clinical trials data. *Drug and alcohol dependence* 2002, 68(2):121-130.

8. Wright C, Sim J: Intention-to-treat approach to data from randomized controlled trials: a sensitivity analysis. *Journal of clinical epidemiology* 2003, 56(9):833-842.

9. Rubin DB: Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics* 1991:1213-1234.

10. Bang H, Davis CE: On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? *Statistics in medicine* 2007, 26(5):954-964.

11. Burton A, Altman DG, Royston P, Holder RL: The design of simulation studies in medical statistics. *Statistics in medicine* 2006, 25(24):4279-4292.

12. Angrist JD, Imbens GW, Rubin DB: Identification of causal effects using instrumental variables. *Journal of the American statistical Association* 1996, 91(434):444-455.

13. McNamee R: Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Statistics in medicine* 2009, 28(21):2639-2652.

14. Atun R, de Jongh T, Secci F, Ohiri K, Adeyi O: A systematic review of the evidence on integration of targeted health interventions into health systems. *Health policy and planning* 2010, 25(1):1-14.

15. Düsing R, Lottermoser K, Mengden T: Compliance with drug therapy—new answers to an old question. *Nephrology Dialysis Transplantation* 2001, 16(7):1317-1321.

16. Dunn G, Maracy M, Tomenson B: Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Statistical Methods in Medical Research* 2005, 14(4):369-395.

17. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH: Instrumental variables: application and limitations. *Epidemiology* 2006, 17(3):260-267.

18. Jo B: Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics* 2002, 27(4):385-409.

19. Little RJ, Long Q, Lin X: A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics* 2009, 65(2):640-649.

20. Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, Wong PK: Medication compliance and persistence: terminology and definitions. *Value in Health* 2008, 11(1):44-47.

21. Dodd S, White IR, Williamson P: Nonadherence to treatment protocol in published randomised controlled trials: a review. *Trials* 2012, 13(1):1-16.

22. Hernán MA, Hernández-Díaz S: Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials* 2012, 9(1):48-55.

Figure 1: The simulation framework
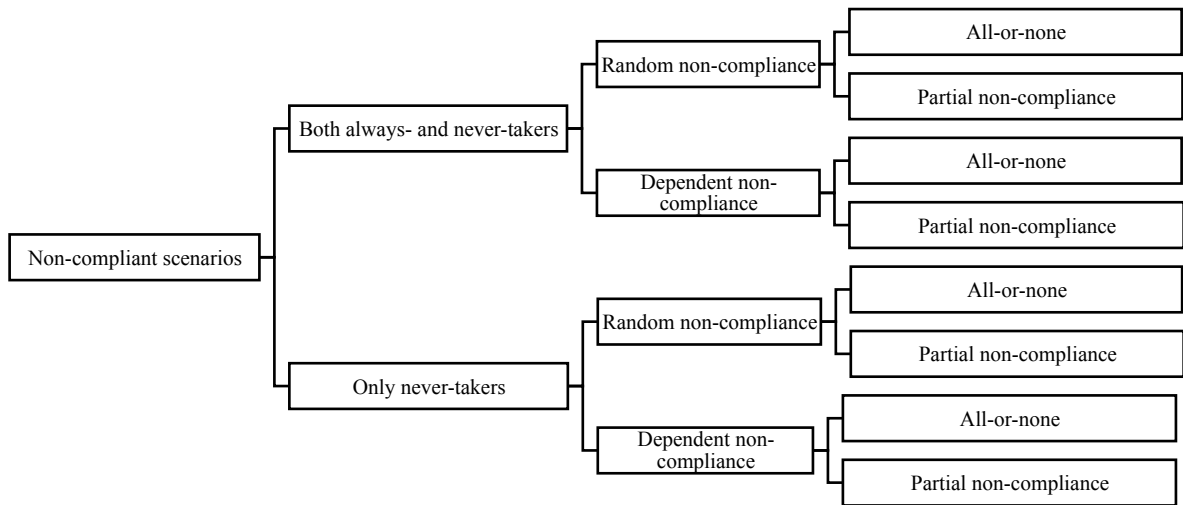
Figure 2: Summary of simulation steps
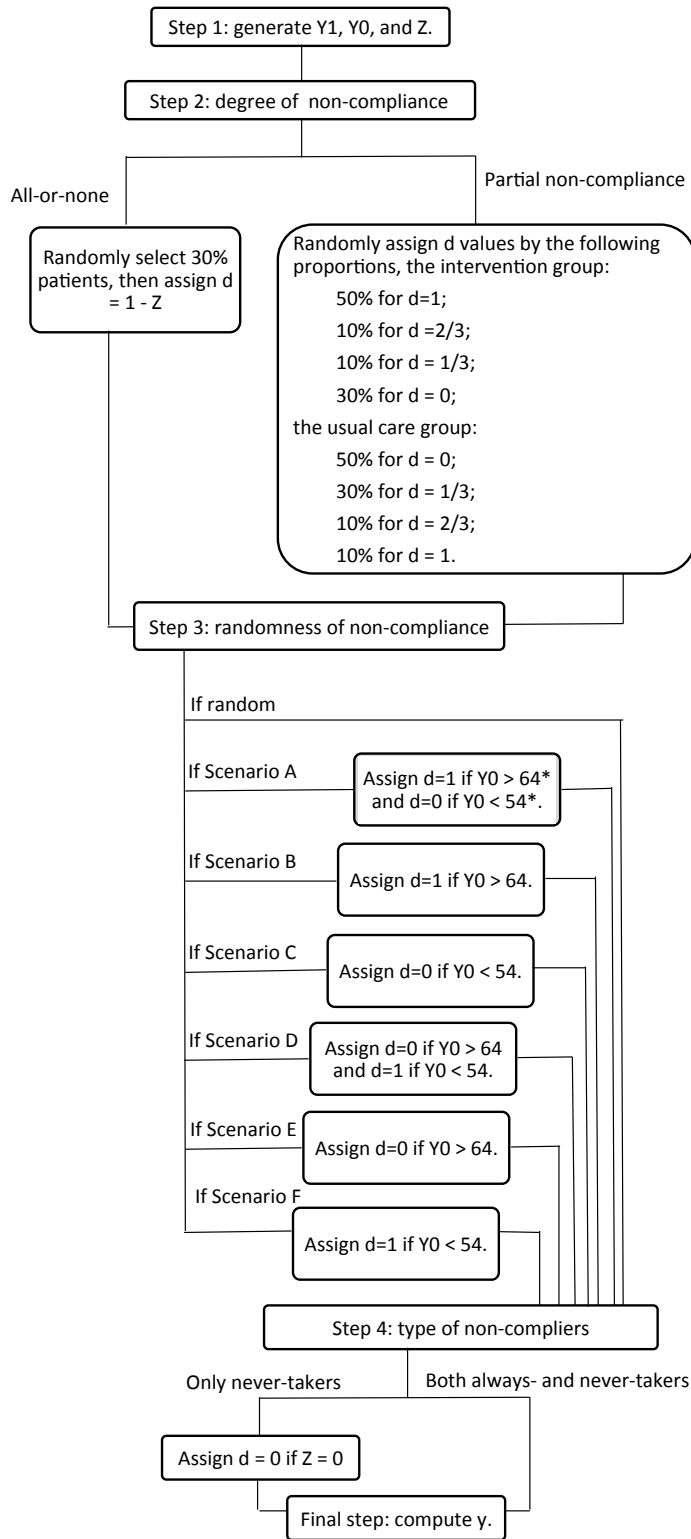
Figure 3: Graphical summary of the bias of estimates under different scenarios (treatment effect = 30)
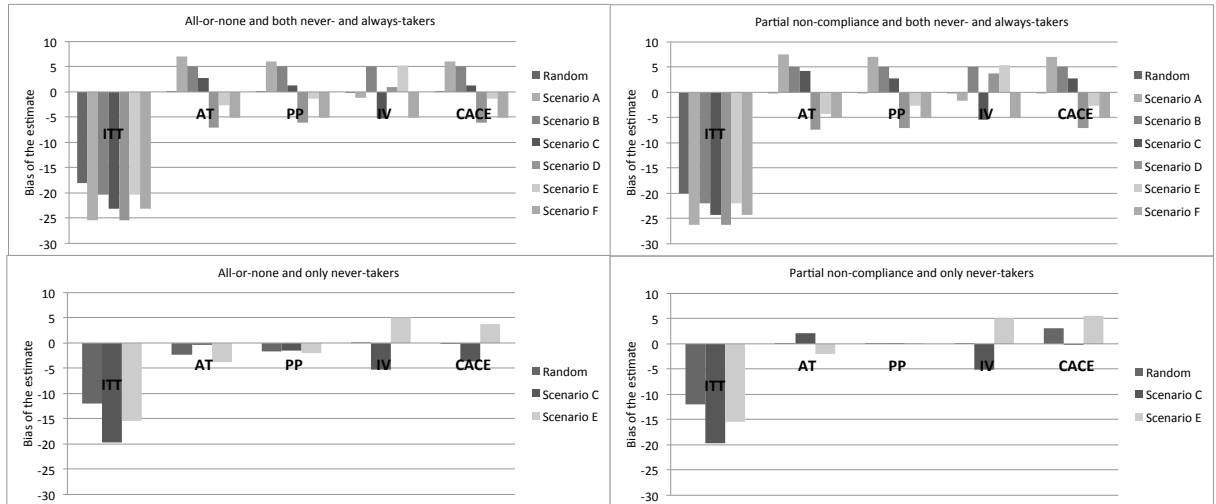
Figure 4: Graphical summary of the bias of estimates under different scenarios (treatment effect = 15)
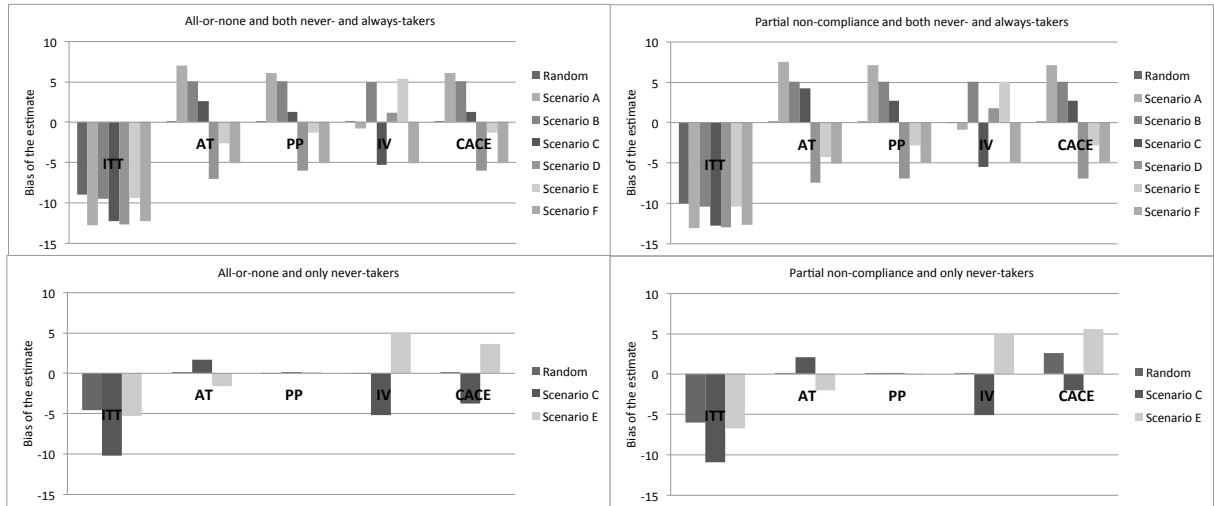
Figure 5: Graphical summary of the bias of estimates under different scenarios (treatment effect = 0)
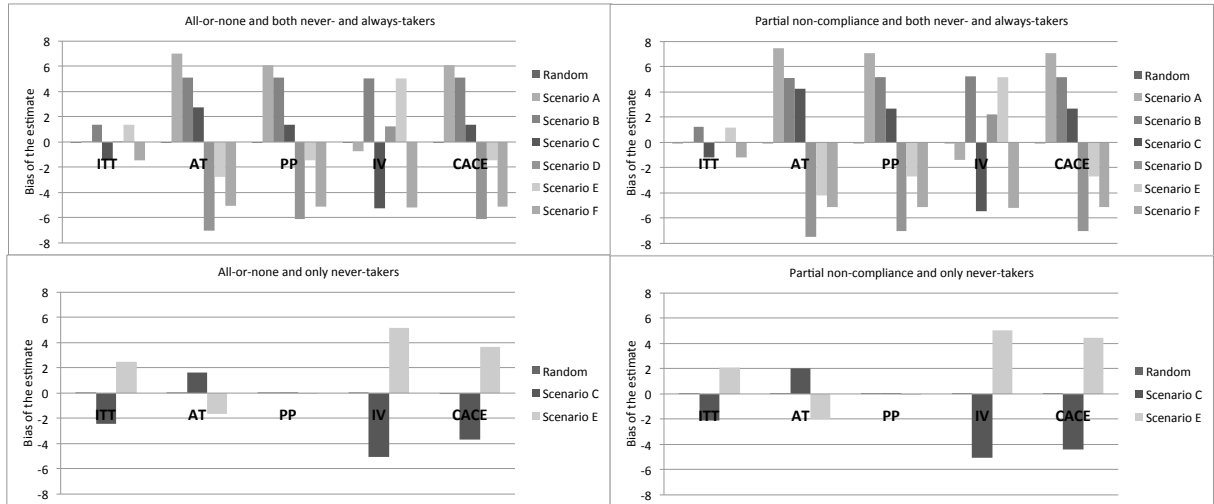
Figure 6: Graphical summary of the mean square error of estimates under different scenarios (treatment effect = 30)
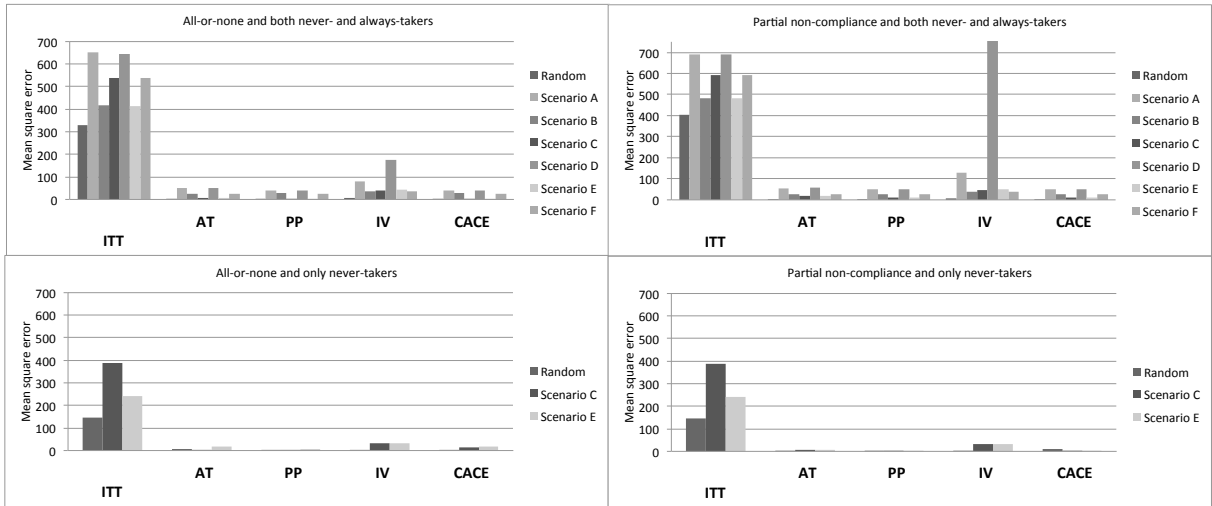
Figure 7: Graphical summary of the mean square error of estimates under different scenarios (treatment effect = 15)
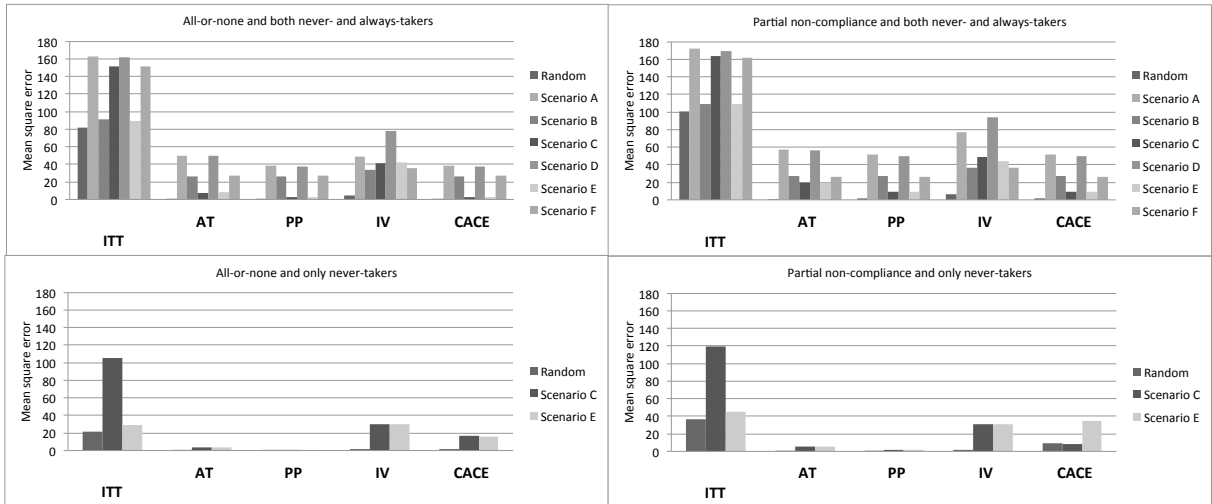
Figure 8: Graphical summary of the mean square error of estimates under different scenarios (treatment effect = 0)
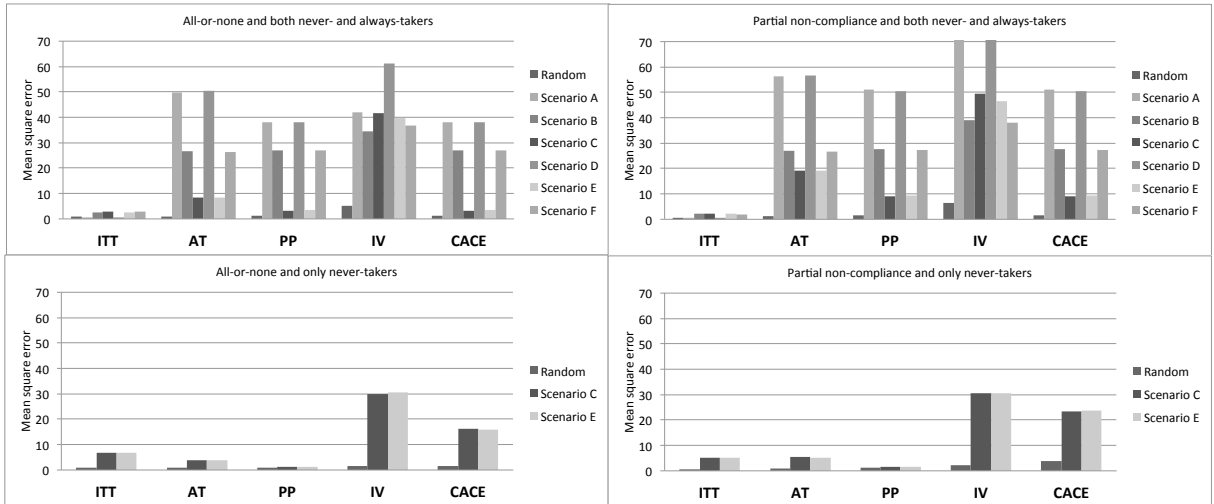
Figure 9: Some recommendations on choosing the analyses of analyzing non-compliant data
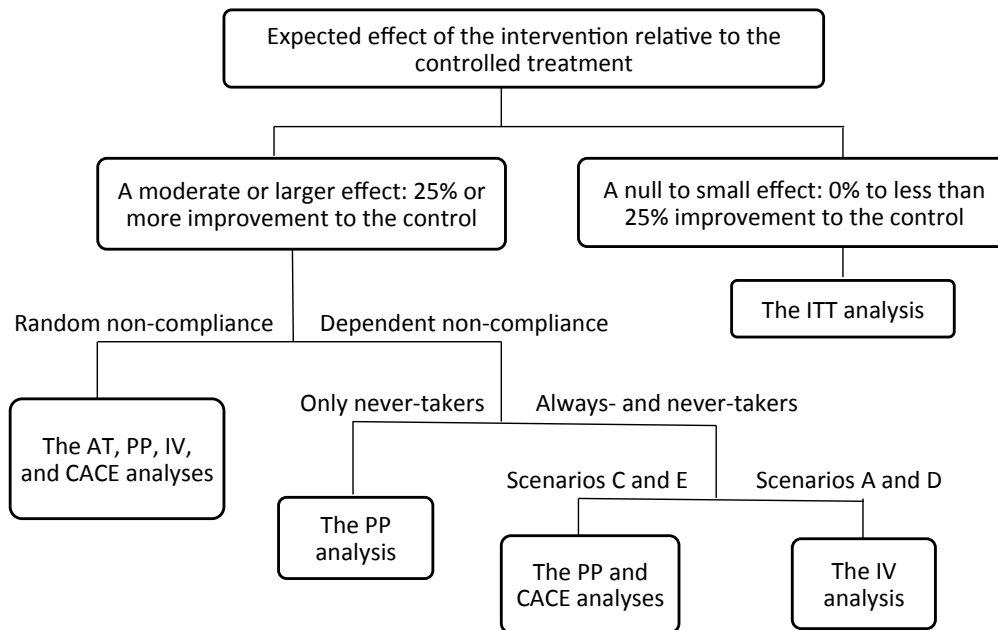
ITT = intention-to-treat; AT = as-treated; PP = per-protocol; IV = instrumental variable; CACE = complier average causal effect
Scenario A: patients with good conditions will always get the intervention and those with poor conditions will reject it.
Scenario C: patients with poor conditions will always reject the intervention.
Scenario D: patients with good conditions will always reject the intervention and those with poor conditions will always get it.
Scenario E: patients with good conditions will always reject the intervention.
Patients who did not receive the intervention would receive the control treatment instead, which was assumed to be the standard care; and there was no other alternative intervention.

Table 1: Summary of the results when both never- and always-takers were allowed (treatment effect = 30)

Scenario	Method	All-or-none				Partial non-compliance			
		Estimate	Bias	MSE	Coverage	Estimate	Bias	MSE	Coverage
Random	ITT	12	-18	330	0%	10	-20	403	0%
	AT	30	0	1	96%	30	0	1	95%
	PP	30	0	1	95%	30	0	2	96%
	IV	30	0	5	96%	30	0	7	95%
	CACE	30	0	1	95%	30	0	2	95%
A	ITT	5	-25	651	0%	4	-26	689	0%
	AT	37	7	50	0%	37	7	56	0%
	PP	36	6	38	0%	37	7	50	0%
	IV	29	-1	80	96%	28	-2	129	96%
	CACE	36	6	38	0%	37	7	50	0%
B	ITT	10	-20	416	0%	8	-22	483	0%
	AT	35	5	27	0%	35	5	27	0%
	PP	35	5	27	0%	35	5	27	1%
	IV	35	5	34	60%	35	5	38	66%
	CACE	35	5	27	0%	35	5	27	0%
C	ITT	7	-23	539	0%	6	-24	590	0%
	AT	33	3	8	18%	34	4	19	2%
	PP	31	1	3	80%	33	3	9	48%
	IV	25	-5	41	73%	25	-5	46	80%
	CACE	31	1	3	79%	33	3	9	47%
D	ITT	5	-25	645	0%	4	-26	690	0%
	AT	23	-7	50	0%	23	-7	56	0%
	PP	24	-6	39	0%	23	-7	51	0%
	IV	31	1	174	97%	34	4	4895	96%
	CACE	24	-6	39	0%	23	-7	51	0%
E	ITT	10	-20	413	0%	8	-22	483	0%
	AT	27	-3	8	18%	26	-4	19	4%
	PP	29	-1	3	78%	27	-3	9	50%
	IV	35	5	42	74%	35	5	49	81%
	CACE	29	-1	3	78%	27	-3	9	48%
F	ITT	7	-23	537	0%	6	-24	591	0%
	AT	25	-5	26	0%	25	-5	27	0%
	PP	25	-5	27	0%	25	-5	27	0%
	IV	25	-5	35	58%	25	-5	39	66%
	CACE	25	-5	27	0%	25	-5	27	0%

MSE=mean square error; ITT=intention-to-treat; AT=as-treated; PP=per-protocol; IV=instrumental variable; CACE=complier average casual effect.

Table 2: Summary of the results when only never-takers were allowed (treatment effect = 30)

Scenario	Method	All-or-none				Partial non-compliance			
		Estimate	Bias	MSE	Coverage	Estimate	Bias	MSE	Coverage
Random	ITT	18	-12	145	0%	18	-12	145	0%
	AT	28	-2	7	36%	30	0	1	94%
	PP	28	-2	4	65%	30	0	1	94%
	IV	30	0	2	95%	30	0	2	95%
	CACE	30	0	2	95%	33	3	12	36%
C	ITT	10	-20	389	0%	10	-20	387	0%
	AT	30	0	1	94%	32	2	5	60%
	PP	29	-1	3	76%	30	0	2	96%
	IV	25	-5	33	35%	25	-5	31	38%
	CACE	26	-4	15	34%	30	0	3	94%
E	ITT	15	-15	239	0%	15	-15	241	0%
	AT	26	-4	16	9%	28	-2	5	63%
	PP	28	-2	6	61%	30	0	2	94%
	IV	35	5	32	41%	35	5	31	39%
	CACE	34	4	16	32%	36	6	4	8%

MSE=mean square error; ITT=intention-to-treat; AT=as-treated; PP=per-protocol; IV=instrumental variable; CACE=complier average casual effect.

Table 3: Summary of the results when both never- and always-takers were allowed (treatment effect = 15)

Scenario	Method	All-or-none				Partial non-compliance			
		Estimate	Bias	MSE	Coverage	Estimate	Bias	MSE	Coverage
Random	ITT	6	-9	82	0%	5	-10	101	0%
	AT	15	0	1	95%	15	0	1	93%
	PP	15	0	1	96%	15	0	2	96%
	IV	15	0	5	96%	15	0	6	96%
	CACE	15	0	1	96%	15	0	2	96%
A	ITT	2	-13	163	0%	2	-13	172	0%
	AT	22	7	50	0%	22	7	57	0%
	PP	21	6	38	0%	22	7	51	0%
	IV	14	-1	49	97%	14	-1	77	97%
	CACE	21	6	38	0%	22	7	51	0%
B	ITT	5	-10	92	0%	5	-10	109	0%
	AT	20	5	26	0%	20	5	27	0%
	PP	20	5	26	0%	20	5	27	0%
	IV	20	5	34	59%	20	5	37	68%
	CACE	20	5	26	0%	20	5	27	0%
C	ITT	3	-12	151	0%	2	-13	164	0%
	AT	18	3	8	23%	19	4	19	3%
	PP	16	1	3	81%	18	3	9	49%
	IV	10	-5	41	72%	10	-5	49	80%
	CACE	16	1	3	80%	18	3	9	49%
D	ITT	2	-13	162	0%	2	-13	169	0%
	AT	8	-7	49	0%	8	-7	57	0%
	PP	9	-6	38	0%	8	-7	50	0%
	IV	16	1	78	96%	17	2	94	97%
	CACE	9	-6	38	0%	8	-7	50	0%
E	ITT	6	-9	90	0%	5	-10	109	0%
	AT	12	-3	8	23%	11	-4	20	3%
	PP	14	-1	3	81%	12	-3	10	45%
	IV	20	5	42	73%	20	5	44	82%
	CACE	14	-1	3	80%	12	-3	10	45%
F	ITT	3	-12	151	0%	2	-13	162	0%
	AT	10	-5	27	0%	10	-5	26	0%
	PP	10	-5	27	0%	10	-5	26	1%
	IV	10	-5	36	60%	10	-5	37	67%
	CACE	10	-5	27	0%	10	-5	26	0%

MSE=mean square error; ITT=intention-to-treat; AT=as-treated; PP=per-protocol; IV=instrumental variable; CACE=complier average casual effect.

Table 4: Summary of the results when only never-takers were allowed (treatment effect = 15)

| Scenario | Method | All-or-none | | | | Partial non-compliance | | | |
		Estimate	Bias	MSE	Coverage	Estimate	Bias	MSE	Coverage
Random	ITT	10	-5	21	0%	9	-6	36	0%
	AT	15	0	1	95%	15	0	1	96%
	PP	15	0	1	95%	15	0	1	97%
	IV	15	0	2	95%	15	0	2	95%
	CACE	15	0	2	95%	18	3	10	62%
C	ITT	5	-10	105	0%	4	-11	119	0%
	AT	17	2	4	66%	17	2	6	58%
	PP	15	0	1	94%	15	0	2	94%
	IV	10	-5	30	24%	10	-5	31	39%
	CACE	11	-4	16	34%	13	-2	8	84%
E	ITT	10	-5	29	0%	8	-7	45	0%
	AT	13	-2	4	68%	13	-2	5	60%
	PP	15	0	1	95%	15	0	2	96%
	IV	20	5	30	24%	20	5	31	36%
	CACE	19	4	16	33%	21	6	34	14%

MSE=mean square error; ITT=intention-to-treat; AT=as-treated; PP=per-protocol; IV=instrumental variable; CACE=complier average casual

Table 5: Summary of the results when both never- and always-takers were allowed (treatment effect = 0)

Scenario	Method	All-or-none				Partial non-compliance			
		Estimate	Bias	MSE	Coverage	Estimate	Bias	MSE	Coverage
Random	ITT	0	0	1	94%	0	0	1	94%
	AT	0	0	1	95%	0	0	1	93%
	PP	0	0	1	95%	0	0	2	94%
	IV	0	0	5	95%	0	0	7	94%
	CACE	0	0	1	94%	0	0	2	94%
A	ITT	0	0	1	95%	0	0	1	95%
	AT	7	7	50	0%	7	7	56	0%
	PP	6	6	38	0%	7	7	51	0%
	IV	-1	-1	42	97%	-1	-1	128	97%
	CACE	6	6	38	0%	7	7	51	0%
B	ITT	1	1	3	59%	1	1	2	67%
	AT	5	5	27	0%	5	5	27	0%
	PP	5	5	27	0%	5	5	28	1%
	IV	5	5	35	58%	5	5	39	65%
	CACE	5	5	27	0%	5	5	28	0%
C	ITT	-1	-1	3	66%	-1	-1	2	72%
	AT	3	3	8	18%	4	4	19	3%
	PP	1	1	3	80%	3	3	9	49%
	IV	-5	-5	42	71%	-5	-5	49	80%
	CACE	1	1	3	79%	3	3	9	47%
D	ITT	0	0	1	95%	0	0	1	95%
	AT	-7	-7	50	0%	-7	-7	57	0%
	PP	-6	-6	38	0%	-7	-7	50	0%
	IV	1	1	61	97%	2	2	544	97%
	CACE	-6	-6	38	0%	-7	-7	50	0%
E	ITT	1	1	3	69%	1	1	2	73%
	AT	-3	-3	8	20%	-4	-4	19	3%
	PP	-1	-1	3	77%	-3	-3	9	48%
	IV	5	5	40	76%	5	5	47	80%
	CACE	-1	-1	3	76%	-3	-3	9	48%
F	ITT	-1	-1	3	59%	-1	-1	2	68%
	AT	-5	-5	26	0%	-5	-5	27	0%
	PP	-5	-5	27	0%	-5	-5	27	1%
	IV	-5	-5	37	58%	-5	-5	38	67%
	CACE	-5	-5	27	0%	-5	-5	27	0%

MSE=mean square error; ITT=intention-to-treat; AT=as-treated; PP=per-protocol; IV=instrumental variable; CACE=complier average casual effect

Table 6: Summary of the results when only never-takers were allowed (treatment effect = 0)

Scenario	Method	All-or-none				Partial non-compliance			
		Estimate	Bias	MSE	Coverage	Estimate	Bias	MSE	Coverage
Random	ITT	0	0	1	95%	0	0	1	96%
	AT	0	0	1	95%	0	0	1	96%
	PP	0	0	1	96%	0	0	1	96%
	IV	0	0	2	95%	0	0	2	96%
	CACE	0	0	2	94%	0	0	4	95%
C	ITT	-2	-2	7	23%	-2	-2	5	35%
	AT	2	2	4	65%	2	2	5	58%
	PP	0	0	1	94%	0	0	2	95%
	IV	-5	-5	30	25%	-5	-5	31	37%
	CACE	-4	-4	16	34%	-4	-4	24	40%
E	ITT	2	2	7	24%	2	2	5	36%
	AT	-2	-2	4	65%	-2	-2	5	60%
	PP	0	0	1	95%	0	0	1	96%
	IV	5	5	31	25%	5	5	30	38%
	CACE	4	4	16	34%	4	4	24	37%

MSE=mean square error; ITT=intention-to-treat; AT=as-treated; PP=per-protocol; IV=instrumental variable; CACE=complier average casual

CHAPTER 5

CONCLUSIONS

There are many statistical and methodological issues in evaluations of integrated care programs. A subset of those issues includes: (1) quantifying the degree of integration for an integrated service network, (2) analyzing randomized controlled trials (RCTs) of evaluating integrated care programs with substantial non-compliance, and (3) assessing the extent of bias incurred when analyzing RCTs of integrated care programs under non-compliance. I have studied those three topics in this sandwich thesis. In this chapter, I will summarize the key results and limitations, and discuss potential implications of the findings on future studies.

In Chapter 2, we proposed a method for quantifying the degree of integration for each service provider in an integrated service network through measuring the agreement of involvement among all members. Each service provider was measured by 4 integration scores, namely, P1, P2, P3, and P4 scores. The P1 score measured the gap between a service provider's level of involvement perceived by the group and the level of involvement expected by the group. The P2 score measured the gap between a service provider's perceived level of involvement and the level of involvement expected by the group. The P3 score measured the gap between a service provider's level of involvement perceived by the group and the level of involvement expected by the provider itself. The P4 score measured the gap between a service provider's perceived level of involvement

and the level of involvement expected by the service provider itself. The P1 score served as the primary measure of integration because it captured the concordance of the whole group towards an agreed-upon integration. We applied the method in measuring the integration of service providers in the Children's Treatment Network (CTN) in Simcoe County and York Region, Ontario Canada, and found generally low scores. By employing spider plots, we identified the service providers who had a large variation across different integration scores. The contrast between integration scores revealed where the gaps in collaboration might be. We showed that the average of all service providers' integration scores was a robust measure of the global integration for the entire network. The results confirmed the integration efforts that had been established one year after the inception of the CTN. At the same time, the findings were important for the CTN planners to identify problems and potential barriers in integrating multiple services.

Our proposed method for measuring integration has some limitations. First, our method primarily measures the service integration and does not consider other types of integration, for example, the functional integration [1]. Second, our method measures the integration by using the responses from representatives of service providers. Respondent bias may exist. Also, the data collected through our method are more likely to reflect the views of the planning group than the frontline teams. The frontline teams may possess different perceived and expected levels of involvement. Finally, our method assumes that responses from all service providers are available. The impact of missing responses on measuring integration needs further investigations.

Despite the limitations, our method has important strengths. It conceptualizes and quantifies the service integration for an integrated service network. Different from the common assumption that a higher level of involvement means a higher integration [2], our method employs a relative score to quantify the gap of integration between perception and expectation. Besides calculating integration scores for individual service providers, our method provides a simple way to graphically display and summarize the global integration score for a whole network. Built on the validated integration framework [3], our method possesses good content relevance and serves as a valid tool for monitoring integration. Studies may also use our method to examine any longitudinal patterns of integration for an integrated service network.

In Chapter 3, we showed that the improvement of psychosocial quality of life among the children receiving the CTN integrated care was not significantly different from those receiving usual care at 24 months. The ITT analysis showed an average improvement of 1.5 points (on a scale from 0 to 100) for the children receiving the CTN integrated care (p value = 0.32). Alternative methods, including the as-treated (AT), per-protocol (PP), and instrumental variable (IV) analyses, did not show a significant effect either but the estimates varied systematically by the magnitude, precision, and direction. All analyses except for the AT analysis showed an estimate that favored the effect of the CTN integrated care. The PP analysis showed the most conservative estimate towards the null effect.

Our results were limited by some methodological challenges for evaluating an integrated care program in a RCT. First, the early development of the CTN integrated care was associated with suboptimal fidelity. It took a longer than expected time to gather local integrated teams and change system management. Thus, the delay of local integrated teams to deliver all planned services for target children partly contributed to high non-compliance in the CTN arm. Second, the usual care group was potentially contaminated by the CTN integrated care. Since all families lived in the same communities, the families assigned in the usual care group might look for services similar to the ones provided by the CTN. On the other hand, employing cluster design RCTs to evaluate large-scale integrated care programs would greatly increase the required resources or sometimes, might not be feasible. Third, the timing of outcome assessment might be too early to show any significant effects of the CTN integrated care. Implementing integration often takes time and the effects on improving chronic conditions also accumulate over time. Finally, the population targeted by the CTN might not be generalizable to a broader population of chronically ill children. Some of those limitations were also noted in the case-control studies that evaluated integrated and community-based care [4]. Integrated care programs are usually designed for patients with multiple and complex health conditions. Most evaluations of those programs are based on non-randomized studies. Developing strategies to deal with those challenges is important to determine any effects of an integrated care program.

In Chapter 4, we assessed the extent of bias in estimating treatment effects by simulating

hypothetical RCTs to mimic the CTN study. We examined 30 non-compliant scenarios

generated by varying three factors: the type of compliers, the randomness of non-

compliance, and the degree of non-compliance. The effect of a hypothetical intervention

was estimated by the ITT, AT, PP, IV, and complier average causal effect (CACE)

analyses. Our findings confirmed that the ITT analysis produced unbiased estimates when

the intervention was not different from the control [5]. However, when an intervention

had a moderate (25% improvement over the control) or large (50%) effect, our results

showed that the ITT analysis was substantially biased. Similar to a previous study [6], we

found that all estimates were unbiased when non-compliance was independent of

patients' outcomes. Additionally, we showed that the PP analysis was unbiased when the

intervention was only accessible to the patients who were offered it. We examined six

non-compliant scenarios dependent on patients' conditions [7] and showed that the IV

analysis was unbiased when non-compliance was symmetrically dependent on patients'

conditions.

Our assessment of bias for common analyses of non-compliance had some limitations.

First, we had not taken into account baseline covariates. Patients' baseline characteristics

might predict their non-compliant behaviours. Second, we assumed a linear relationship

between the effect of an intervention and the degree of non-compliance. The linear

assumption might not be reasonable for certain interventions, for example, the

interventions whose effects were only dependent on some key components. Third, we did

not consider missing data in the simulations. Patients' follow-up outcomes would probably not be available if they withdrew from the trials. Finally, our simulation only simulated a subset of non-compliant scenarios and the findings might not be generalizable to more complex cases. Future investigations by overcoming those limitations will provide further insights into the issue of dealing with non-compliance in RCTs.

In summary, this sandwich thesis investigated three issues in the evaluation of integrated care programs. Our proposed method to quantify integration for an integrated service network was carefully described and illustrated. By conducting different analyses to handle substantial non-compliance, we did not find a significant difference between the CTN integrated care and usual care on target children's psychosocial function at 2 years. Our results also showed that propensity score methods, which had been mostly used in observational studies, could be used to adjust for potential confounding in AT and PP analyses. Methodological challenges identified from the CTN trial informed the design of future studies to evaluate similar integrated care programs in a RCT setting. Through generating hypothetical RCTs based on the CTN study, we showed that the ITT analysis under non-compliance was substantially biased when the intervention had a moderate or large effect relative to the control. Our simulation study also contributed to the literature by showing that the PP analysis was mostly unbiased when only assigned patients had the access to the new intervention with moderate or large effect. Based on the results, recommendations for common analyses were made to provide unbiased or less biased estimates under different non-compliant scenarios.

References

1. Gillies RR, Shortell SM, Anderson DA, et al. Conceptualizing and measuring integration: findings from the Health Systems Integration Study. Hospital & Health Services Administration 1993;38(4):467–89.

2. Ahgren B, Axelsson R. Evaluating integrated health care: a model for measurement. International Journal of Integrated Care [serial online] 2005 Aug 31; 5. Available from: http://www.ijic.org/index.php/ijic/article/view/134/267. URN:NBN:NL:UI:10-1-100376. Assessed 21 May 2012.

3. Browne G, Roberts J, Gafni A, et al. Conceptualizing and validating the human services integration measure. International Journal of Integrated Care [serial online] 2004 May 19; 4. Available from: http://www.ijic.org/index.php/ijic/article/view/98/195. URN:NBN:NL:UI:10-1-100340. Assessed 21 May 2012.

4. Bardsley M, Steventon A, Smith J, et al. Evaluating integrated and community-based care: How do we know what works? Nuffield Trust, Research summary 2013. http://www.housinglin.org.uk/_library/Resources/Housing/OtherOrganisation/Nuffield_evaluating_summary.pdf. Assessed 26 January 2014.

5. Hernán MA, Hernández-Díaz S: Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials* 2012, 9(1):48-55.

6. Bang H, Davis CE: On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? *Statistics in medicine* 2007, 26(5):954-964.

7. McNamee R: Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Statistics in medicine* 2009, 28(21):2639-2652.