Economic Foundations of Theories of Change

May 3 & 4, 2021

Canadian Evaluation Society 2021

May 4

Gregory Mason

University of Manitoba


TRADITIONAL TERRITORIES
ACKNOWLEDGEMENT
The University of Manitoba campuses are located on original lands of Anishinaabeg, Cree, Oji-Cree, Dakota, and Dene peoples, and on the homeland of the Métis Nation.
We respect the Treaties that were made on these territories, we acknowledge the harms and mistakes of the past, and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of reconciliation and collaboration.
umanitoba.ca
University of Manitoba

Plan for the workshop

Day 1

Hour 1: Key themes from Economics

Hour 2: The Policy Map: An economist's view

Hour 3: Applying core ideas from economics the 2021 federal budget – Breakout 1

Day 2

Hour: 1 Modelling policy: Cause and effect

Hour: 2 Measuring economic impact – Evaluating a basic income

Hour: 3 Using economics to design an – Breakout 2

Note: all images/videos in this presentation have been acquired under licence, creative commons, or lie in the public domain (YouTube) for the "live" presentation.

Breakouts

- Each day will have a breakout (start of Hour 3) with follow-up to conclude the Day.
- I have preassigned you to a breakout …. very imaginatively entitled 1, 2, and 3.
- The email you received will have identified a policy/program and associated issues/questions for each breakout.

Modelling Policy – Cause and effect

What are good explanations?

- Good explanations use causal processes (mechanisms) that link initial conditions, changes in certain "state" variables, and the resulting changes in social reality.
- Observing "regularity" (correlation) between the change in state and the change in social reality is not enough.
- To explain variation in the measured outcomes, we need to know the "why"

the mechanism

- The mechanism links output and outcome
- "Why" and "how" come from a theory of change.
- Confirming the mechanism requires *direct* or *indirect* manipulation of inputs to assess changes in outcomes.

Cause and effect (A very basic explanation)

Necessary causes:

- For *X* to be a necessary cause of *Y*, then if *Y* occurs, *X* must also occur. The fact that *X* occurs does not imply that *Y* will occur.

Sufficient causes:

- For *X* to be a sufficient cause of *Y*, then the presence of *X* always implies that *Y* will occur. The fact that *Y* occurs does not imply that *X* has occurred, since another variable (*Z*) could be the cause.

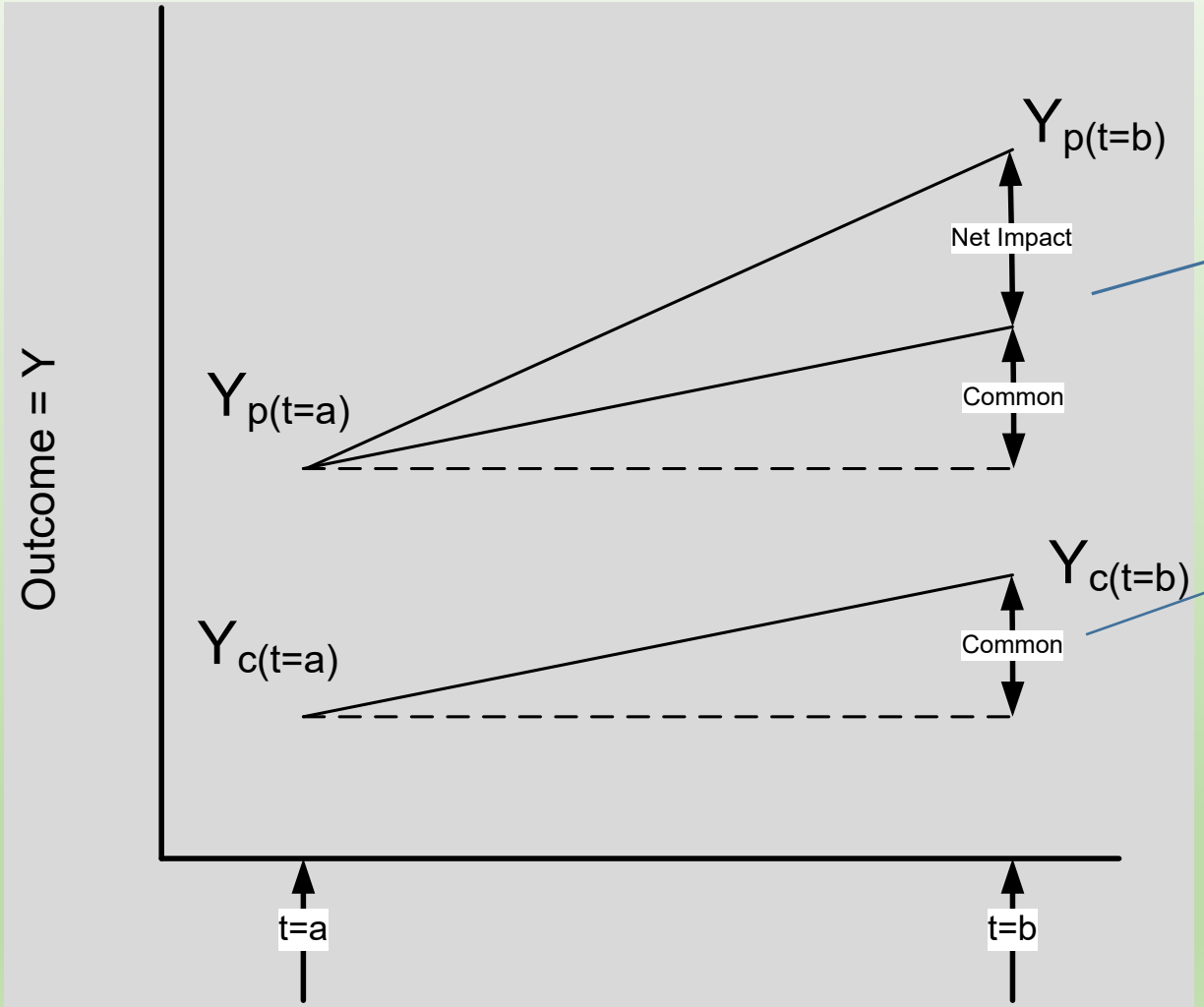
Contributory causes:

- A cause *X* may *contribute* to the occurrence of *Y*, if *X* occurs before *Y* and varying *X* varies *Y*.

Causal glossary

- **Independent (exogenous, cause) variables** – are the direct policy/program interventions and socio-economic control.
- **Dependent (endogenous, effect) variables** – represent the outcomes.
- **Intervention variables** – special class of independent variables that represent policy/programming.
- **Discrete (dummy, 0-1) variable** – marks the "boundary" between the program and counterfactual or expresses binary states.
- **Counterfactual** – the state of affairs that would have occurred without the program/intervention
- **Gross impact** – observed total change in the outcome(s).
- **Net impact** – portion of gross impact attributable to the program intervention.
- **Experiment** – the purposeful manipulation of independent and intervention variables to observe the change in outcomes.
- Randomized control trial (RCT) – subjects enter a treatment or control group based on random selection.
- **Quasi-experiment** – the replication of manipulation within the context of a statistical model.
- **Natural Experiment** – using real world implementation to mimic a lab experiment.
- **Necessary condition** – the "cause" (output) is essential to the outcome.
- **Sufficient condition** – the "cause" (output) will always be associated with the outcome and by itself will produce the outcome.
- **Contribution** – the "cause" output contributed to the outcome, by may be neither necessary nor sufficient and represents a confluence of theory and expert judgement.

The meaning of net impact



Results for the treatment group

Results for the comparison group

- We observe the program group outcomes Yp at time t=a and t=b. We also observe the control/comparison group outcomes Yc t=a and
- The control/comparison group reflects the forces common to the program and comparison group.
- The net impact is the increment (or decrement) uniquely attributed to the program.

The Difference in Differences (DID) estimator uses the average before and after values for an outcome variable for the program and comparison group.

$$DID = [Y_p\ (t=a) - Y_p(t=b)\ ] - [Y_c(t=a) - Yc(t=b)]$$

Example:

Avg. earned income before (program group) = $4500
Avg. earned income after (program group) = $6500
Avg. earned income before (comparison group) = $10,500
Avg. earned income after (comparison group) = $11,000

DID = [6500 – 4500] – [11,000 – 10,500] = $1,500
= net impact attributable to the program (treatment)

The concept of a net impact is fundamental to the evaluation of programs and policies.

It is a hallmark of economic evaluation

Limits of randomized designs

Randomized double blind experiments (RCTs) are cited as the "Gold Standard"

But since these are usually not feasible in social-economic-environmental policy (SEEP), this is a misleading reference point.

RCTs do not work for economics because

- Human subjects are unreliable (they move, die, or otherwise fail to participate in the full experiment).
- Many see the administration of a placebo (such as offering a subsidy) as withholding a treatment.
- Social/economic/environmental policy cannot be masked (creating a placebo is difficult).

Causal inference in SEEP

- Thought experiment
- Natural experiment
- Quasi-experiment
  - Pre-post
  - Multivariate regression
  - Statistical matching

Thought Experiment

- Theory of change expressed as algebraic/graphical model
- Analysis to create a range of outcomes under alternate assumptions
- Made popular by Einstein, but common to economics where analytic models support manipulation to derive predictions under a set of assumptions (the shifting of demand and supply to examine impacts of taxes, price floors/ceilings, etc. are examples)
- Includes scenario design (simulation) using software (Excel and other simulation software)
- Used to design a program, scope the range of outcomes in advance of evaluation, and prepare “synthetic” outcomes where empirical data are sparse
- If the model parameters are speculative (empirically weak) these experiments can mislead.

The various climate change models and the infection projections of COVID-19 are typical examples.

Natural Experiments

Natural experiments have emerged as a major technique for inferring causation from observational data. (see FP Op ed)

Basic method

- Create/identify a "split" in the sample, where treated and untreated are classified by a variable that is not related to the treatment.
- This creates a "treatment" and "control"
- This split often occurs "naturally" where the program change occurs in one area/jurisdiction or over time and not in others that are "closely similar."
- Pre-post analyses are inherently weaker than natural experiments that are coterminous.
- Now, make the argument that the treatment and control group are "close" in all variables except for the existence of the treatment
- Difference-in-differences (DID) methods are a common evaluation framework.

Quasi Experiments

Quasi-experiments have three critical features.

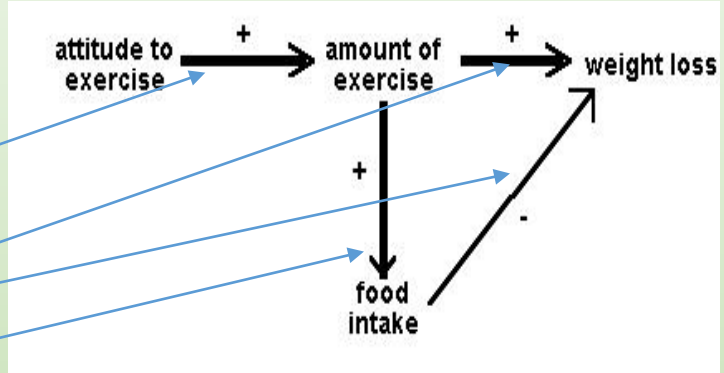
- The information available allows the researcher to define two mutually exclusive subsets:
  - *a program group* consisting of observations (respondents) who participated in the program
  - *a comparison group* consisting of observations (respondents) who are otherwise identical to the program group, along observed characteristics, except for program participation
- The data spans the period before and after when policy intervention applies.
- All observations have a range of attributes that act as covariates.

For an *RCT*, the research need only know whether an observation is associated with the treatment or control, and a common outcome, believed to be affected by the intervention.

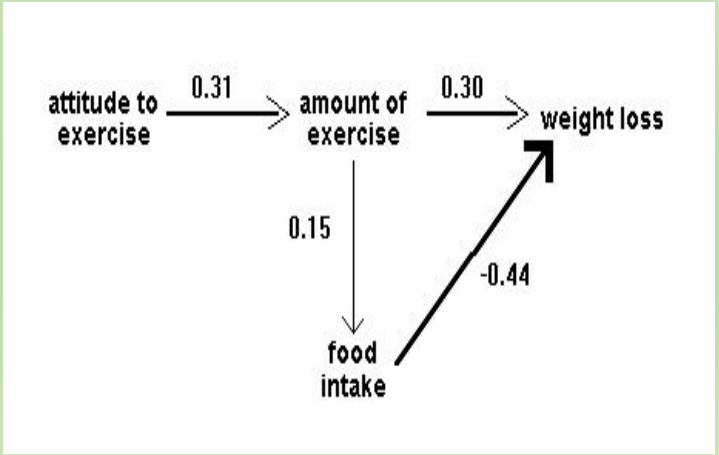
A *quasi experiment* has the same requirement, plus information on the attributes of each observation (age, income, gender, etc.)

Path analysis

- Path analysis is multiple regression applied to various elements of a theory of change
- Imagine we were interested in how food intake creates weight gain.
- Input path
- Output path

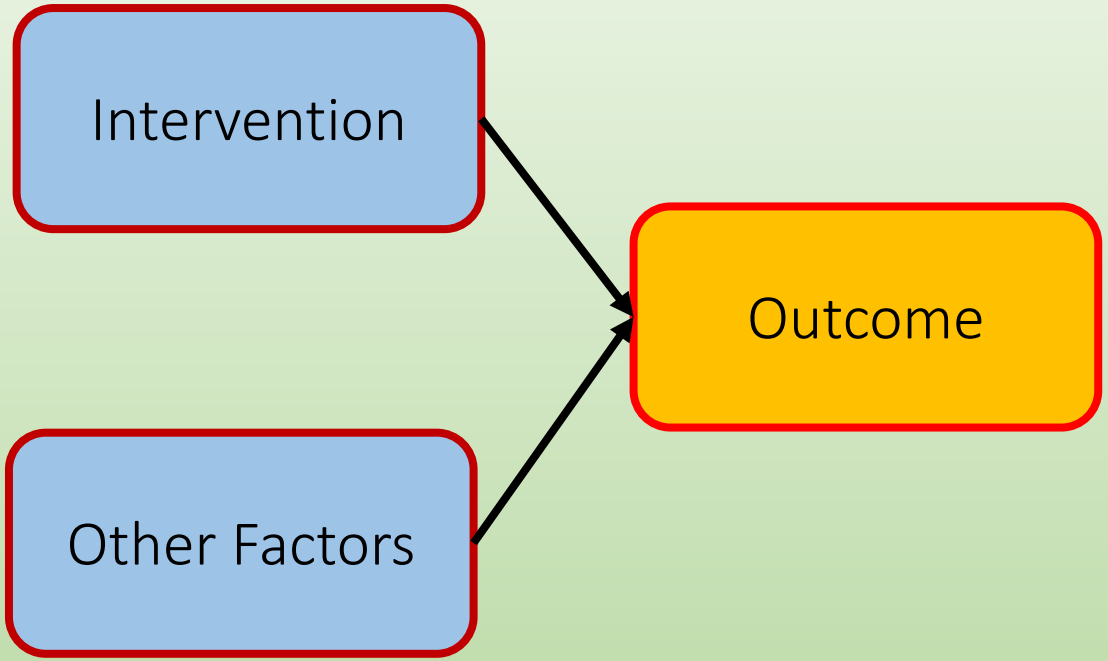


Shows the theoretical determinants of exercise and food intake



Shows the empirical determinants of weight loss/gain due to food intake

Returning to causal logic models

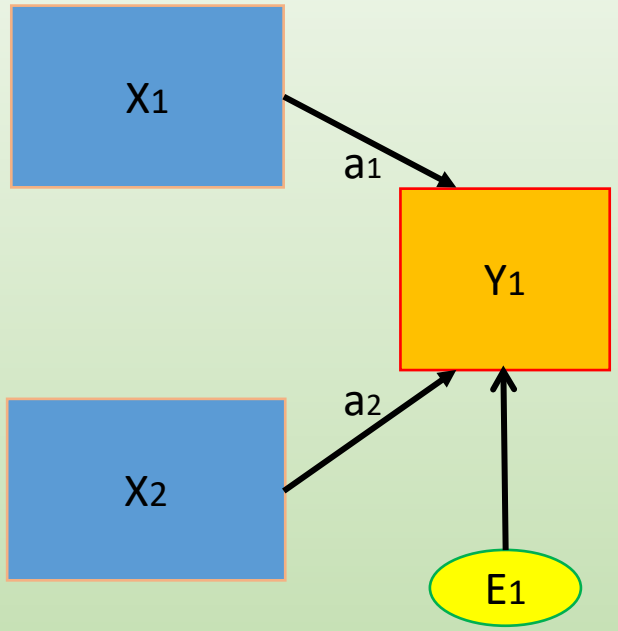


The causal logic model clarifies the theory of how interventions produce outcomes.

Multiple methods and experimental techniques establish the relative importance of causes of changes in outcomes

The measured outcome must be a reliable and a valid measure of a program/policy goal

Causal Analysis

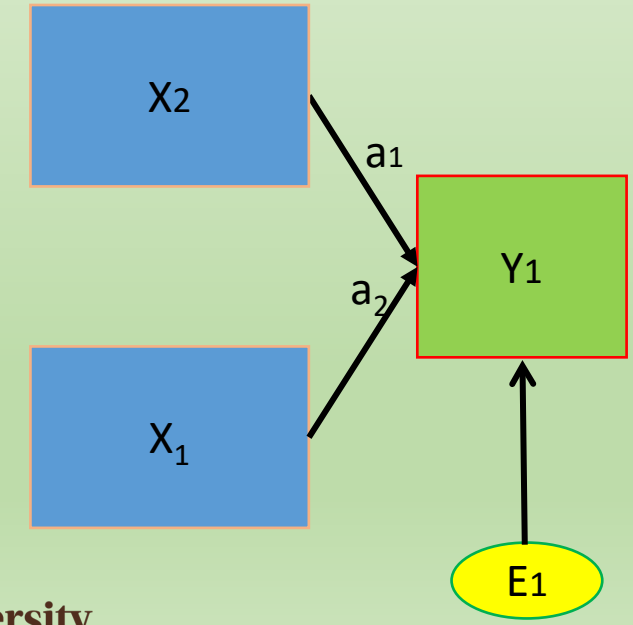


- X_1, X_2 are independent (causal) variables also known as exogenous variables.
- Y_1 is a dependent (effect) or endogenous variable.
- e_1 is an error term, reflecting measurement imprecision, poor model design, failure to include all the relevant variables, external factors, etc.

$$Y_1 = a_0 + a_1X_1 + a_2X_2 + e_1$$

Regression as a quasi experiment

- "Treatment effect" is the effect of a given treatment or intervention on an outcome variable of interest.
- In the simple regression model



$Y_1 = a_0 + a_1X_1 + a_2X_2 + e_1$

where X1 is the 0 -1 treatment variable, the treatment effect is the coefficient a1.

a_2 measures the effect other factors

Adding more covariates will usually weaken the importance of a_1

For each observation $X_1 = 0$ for no treatment and $X_1 = 1$ for the treatment. (it makes no statistical difference which is designated as 1 or 0)

A pre-post design defines $X_1 = 0$ as observations before the intervention and $X_1 = 1$ as observations after the intervention

The coefficient a1 measures the strength of the effect of the intervention

University of Manitoba

Measuring economic impact – Evaluating a basic income

Basic income has two variants

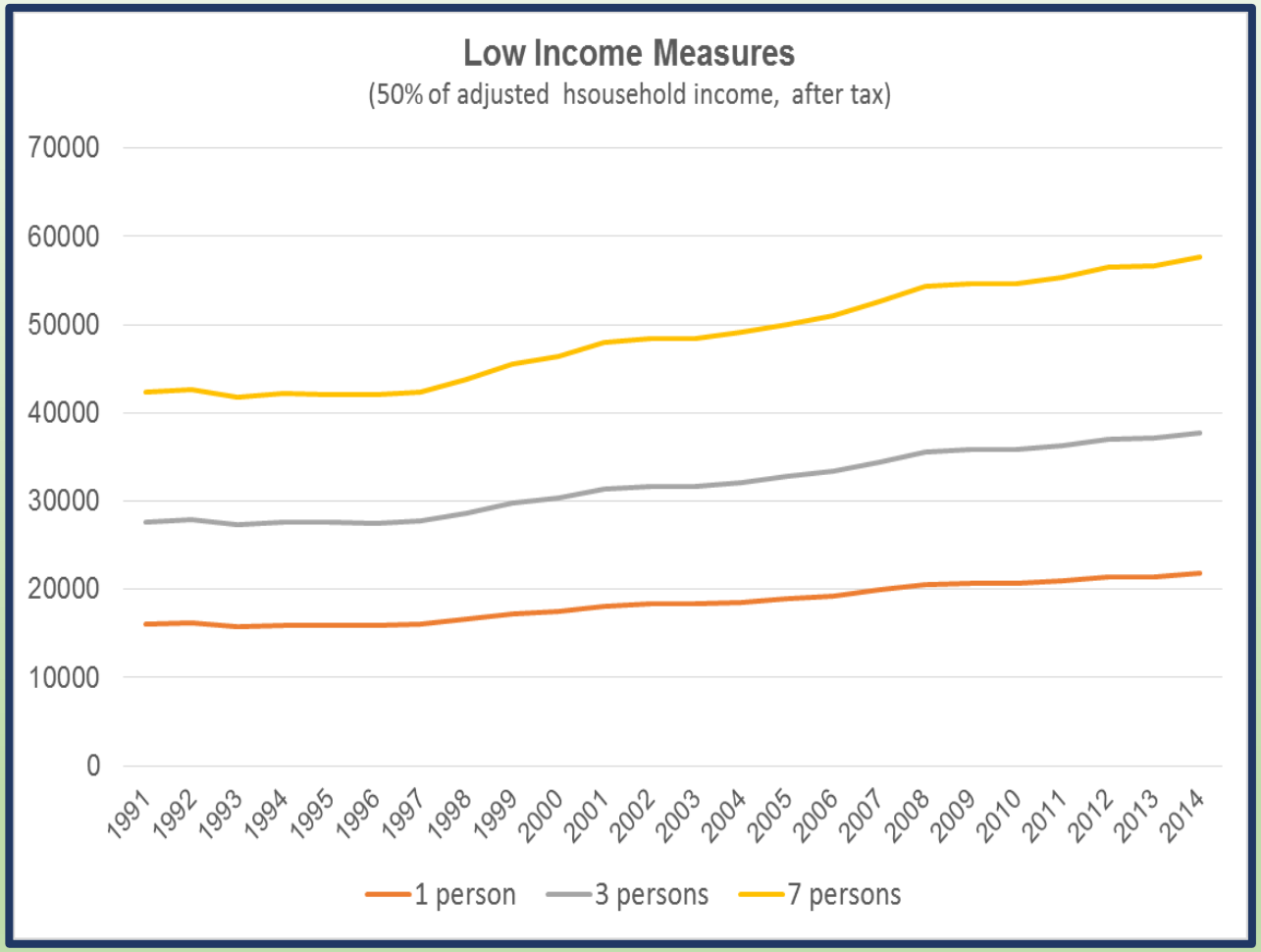
- Universal basic income (UBI) where everyone is receives a minimum income.
- Negative income tax (NIT) that guarantees a minimum, and increases support as a reduced rate as earnings rise.

What does an ideal basic income look like?

- Enables individuals to have both
  - (1) autonomous income to meet their needs; and
  - (2) access to public services that benefit all of us;
- Replaces income provided through social assistance systems and other supports such as GST rebate;
- May not replace other income support such as Employment Insurance, old age security and public pension
- Is inflation adjusted and declines as other income increases
- Can be adjusted to meet specific needs (lone parenthood) ,
- Makes no one worse off by the transition from the existing system.
- Does not negate the need for labour adjustment programs or education
- Does not eliminate the income tax system

Adapted from the Basic Income Canada Network

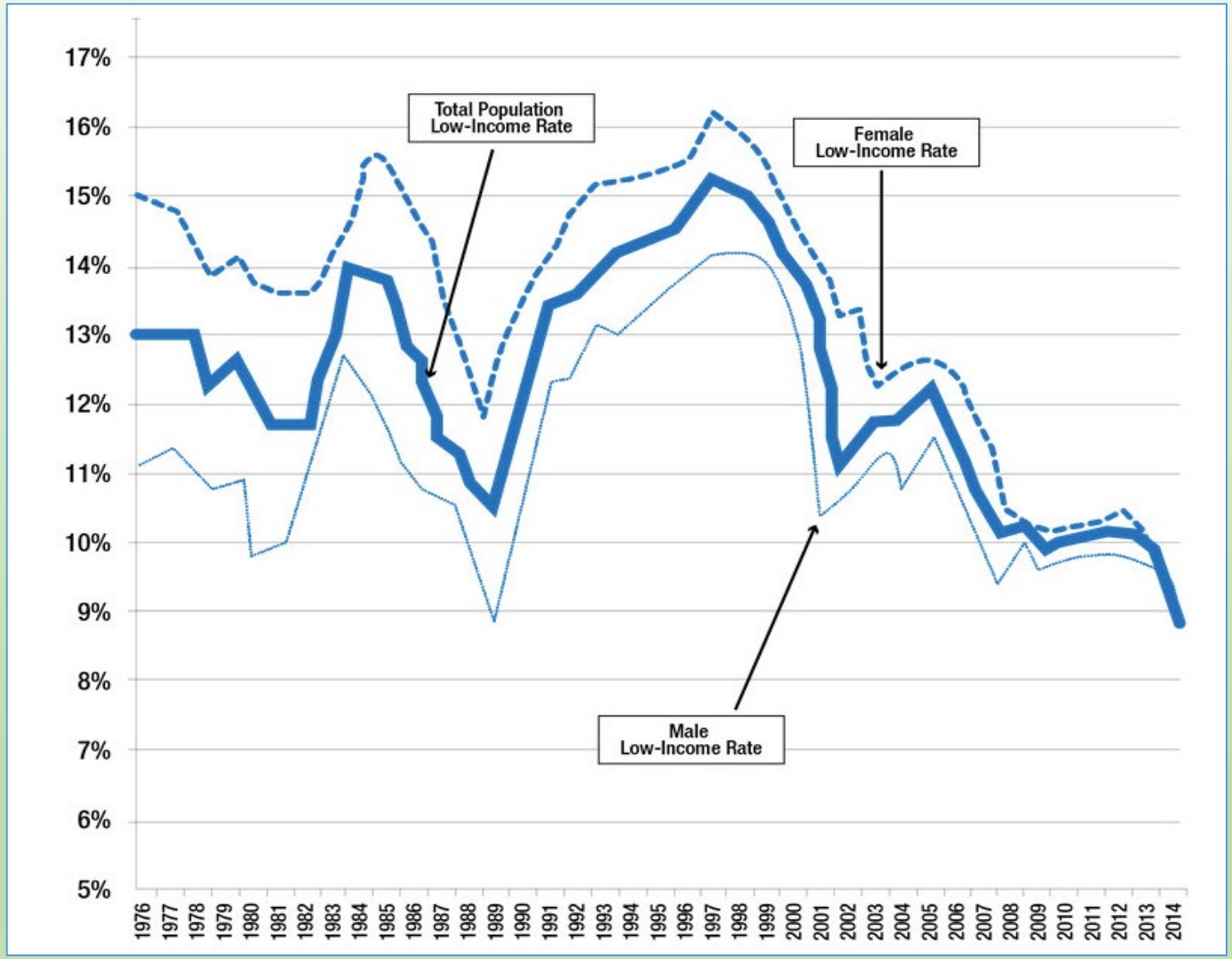
Reality Check – poverty and inequality in Canada



In 2014 a single person would be judged as "poor" if their income were below $21,773

The Parliamentary Budget Office has costed the basic income for Canada at about $75 billion per year. If all other support to low income persons is also cancelled, the net cost to the federal government is closer to $42 billion or about 14% of the total budget

Canada's poverty rate has declined since 1997



Poverty has declined in Canada over the last 30 years.

The gap between male and female poverty rates has narrowed and all but disappeared.

Poverty among Canadian seniors has declined very sharply is now less than Canada's poverty rate for all persons.

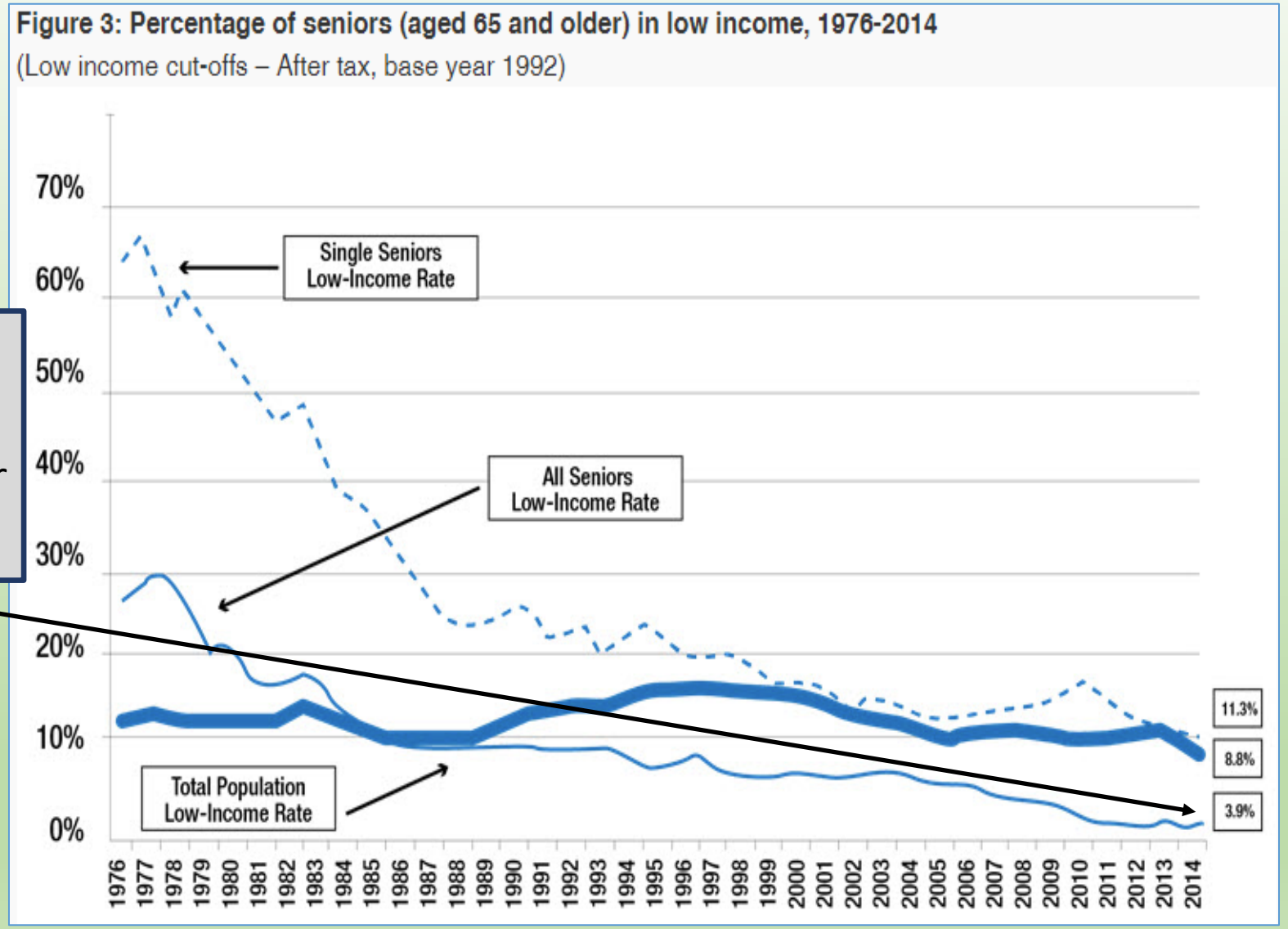
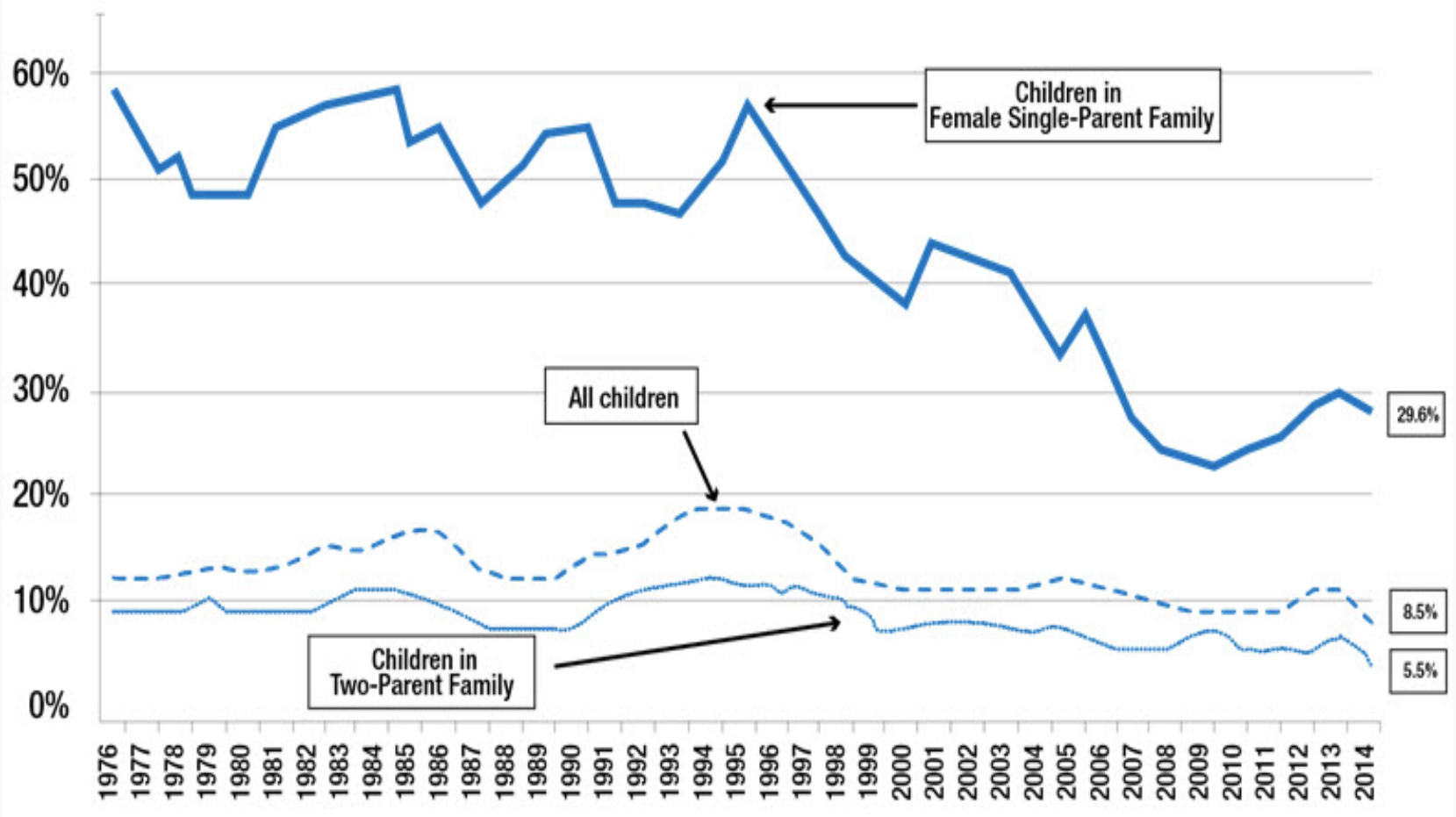


Figure 6: Trends in low-income rates of children (under 18 years old) by family type, 1976-2014

(Low income cut-offs – after tax, base year 1992)



Children in female single-parent families are at most risk of experiencing poverty.

Children in two parent families experience much less poverty than those in single parent families.

The Manitoba Basic Annual Income Experiment (Mincome)

1974 – 78 (operations phase)

1981 – p (research phase)

Simpson, W., Mason, G. and Godwin, R. (2017) "The Manitoba Basic Annual Income Experiment: Lessons Learned 40 Years Later" *Canadian Public Policy*, Volume 43, Number 1, March/mars 2017, pp. 85-104 https://doi.org/10.3138/cpp.2016-082

What was (is) unique about Mincome?

- Mincome was and remains a landmark micro-economic longitudinal study.
- It preceded development of the SWH/LMAS/ SLID surveys of work history and the long-form census.
- It paralleled other major income maintenance experiments in the United States.

A focus on labour market behaviour conditioned all aspects of Mincome design

Mincome had two primary and one secondary goal,

1. To ‘‘evaluate the economic and social consequences of an alternative social welfare system based on the concept of a negative income tax”… and
2. To ‘‘examine the labour supply responses of households and individuals to a guaranteed annual income’’
3. Over time, a third not explicitly stated in the design documents, evolved, which was to understand the administrative and logistical challenges involved in implementing such a system across the population.

Mincome was a longitudinal RCT

Research Questions

Mincome focused on testing a single null hypothesis "*a NIT will not reduce labour market attachment*"

All household members over 15 provided details on their labour force participation

Other data collected served as mediators for testing this hypothesis

Over time, researchers attempted to include other research objectives (such as impact on wealth, family relationships…) but these were always seen as secondary to the main purpose of testing labour force attachment.

Structure

Three sites supported the analysis

- Winnipeg (Main site) – Full RCT
- Dauphin – Single plan, volunteer sample
- Rural dispersed – Single plan, random selection

Dauphin is termed a "saturation" site, where all residents with qualifying income and wealth could enrol. Surprisingly, many who were eligible chose not to participate. (over 40%)

The Winnipeg Sample (n=526)

Table I: Mincome Guarantee Levels and Tax Rates by Plan Type (Just Before Periodic Survey 1): Main Winnipeg Sample

	Tax Rate on Total Income		
Guarantee at enrolment, $	35%	50%	75%
3,800	Plan 1 ($n = 55$)	Plan 3 ($n = 61$)	Plan 6 ($n = 49$)
4,600	Plan 2 ($n = 67$)	Plan 4 ($n = 70$)	Plan 7 ($n = 29$)
5,400	Plan X	Plan 5 ($n = 56$)	Plan 8 ($n = 45$)
		Plan 9 (Control; $n = 94$)	

Mincome allocated participants to one of these nine plans using a complex sample design (Watts-Conlisk Model)

A birds-eye view of Mincome

The Baseline survey collected data in 1974/75 on earnings, work history, income, etc. to identify admissible low-income families and determine initial payments

Periodic surveys, approximately every 3 – 4 months tracked labour market, income, and other changes for the household

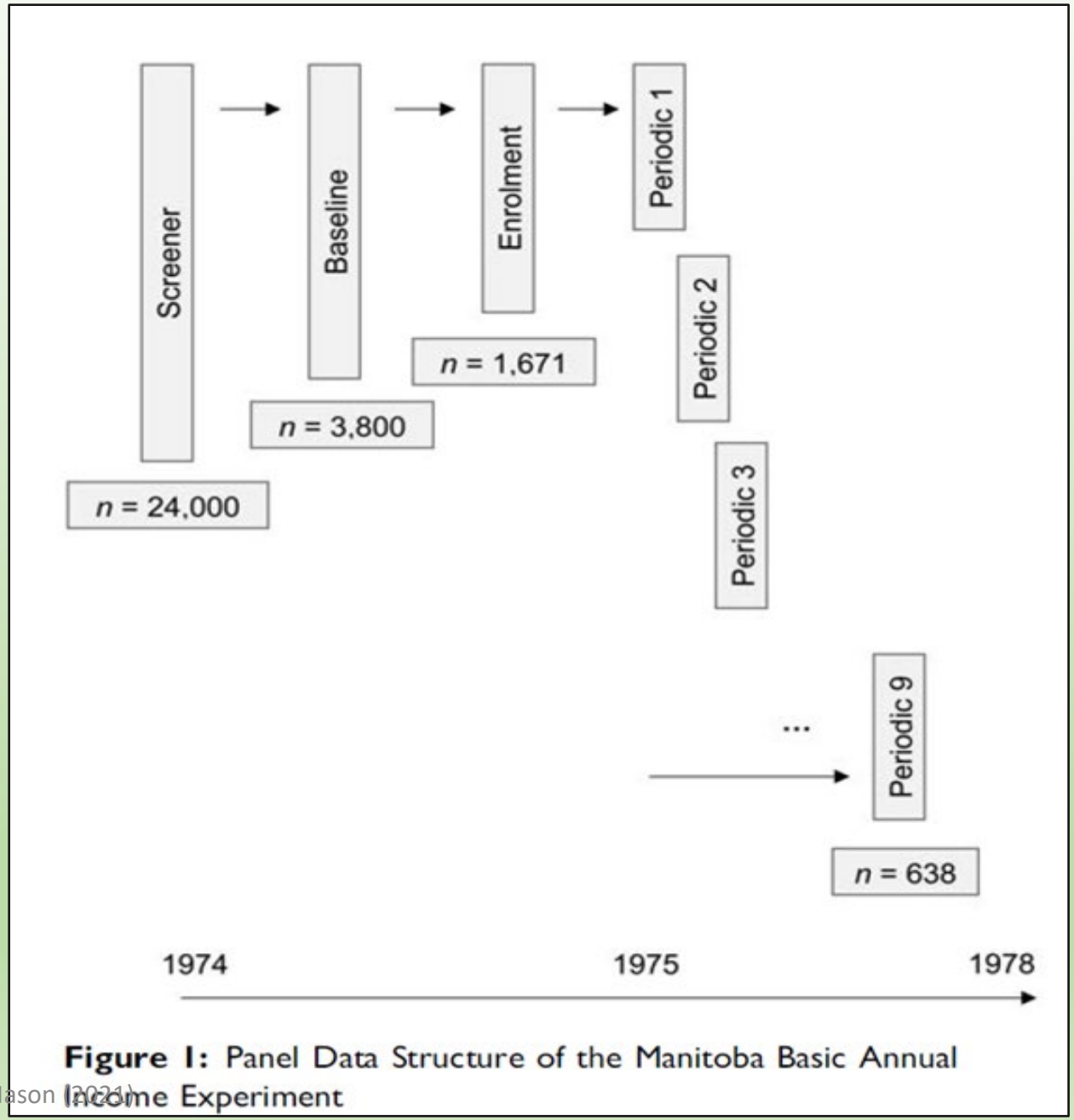
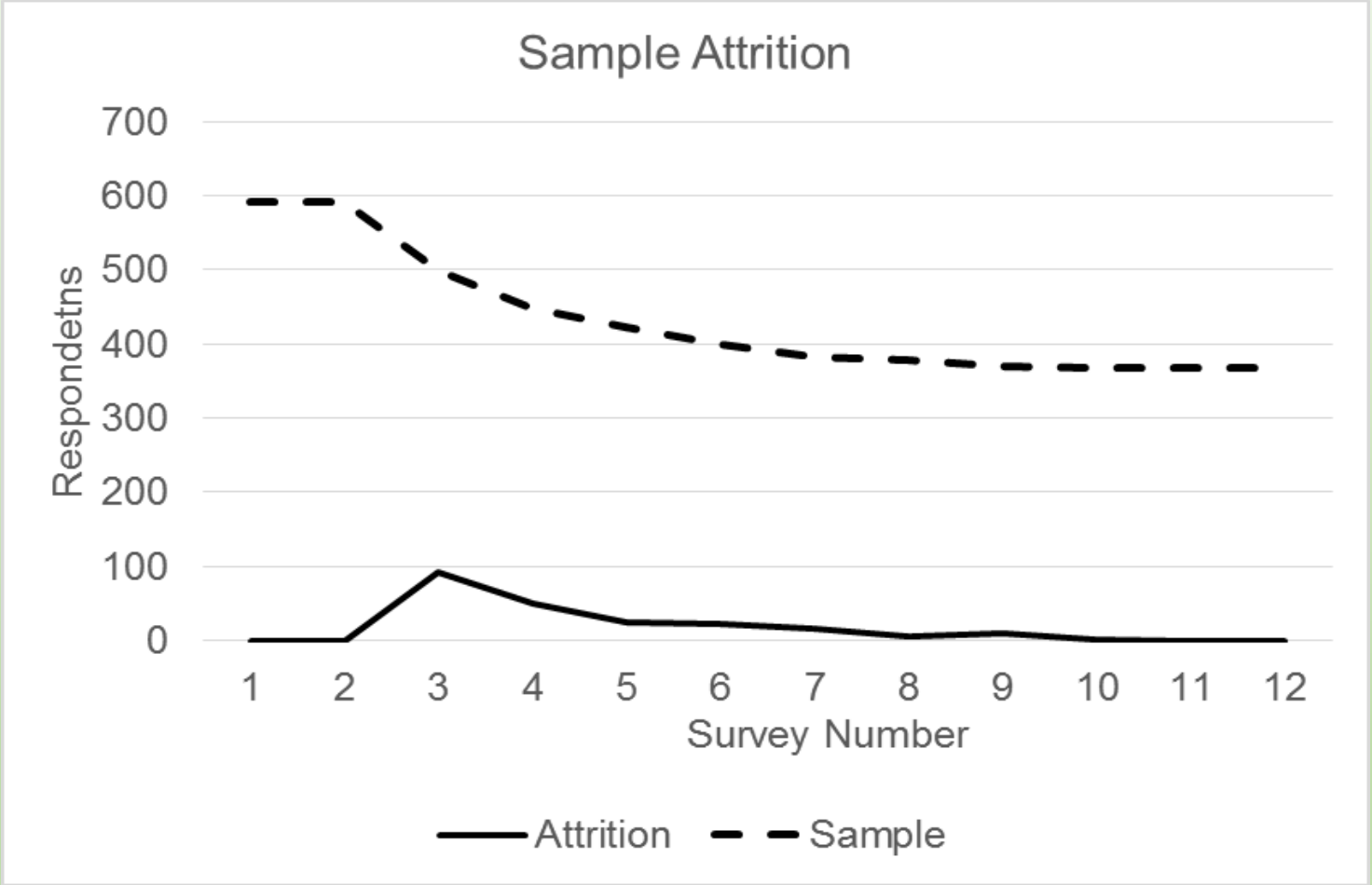


Figure 1: Panel Data Structure of the Manitoba Basic Annual Income Experiment

Survey based longitudinal data's "dirty secret" - Attrition

Mincome samples shrink by about 30% during the panel period

Participants left the experiment for many reasons, but often because their earnings rendered them ineligible for benefits



Mincome Findings

- Mincome/U.S. experiments were designed to answer questions about the work disincentive effects of a negative income tax (guaranteed income) program
- Dispersed Winnipeg sample linked family labour supply response and other data to randomly allocated (?) treatments (differing guarantees and tax rates) plus a control group
- Allow direct estimates of experimental impact
- Modest reduction in work effort …1% for men, 3% for wives and 5% for unmarried women

Quasi-experimental findings from Mincome

- Administrative health records for Dauphin estimate the impact of Mincome on health behaviour and outcomes.
- This study did not use Mincome data, but aligned health data to Dauphin and developed a matched control group elsewhere in MB.
- Quasi-experimental estimates find that Mincome's "town without poverty" reduced hospitalization rates by 8.5%
- Reduction in
  - accidents and injuries
  - mental health diagnoses

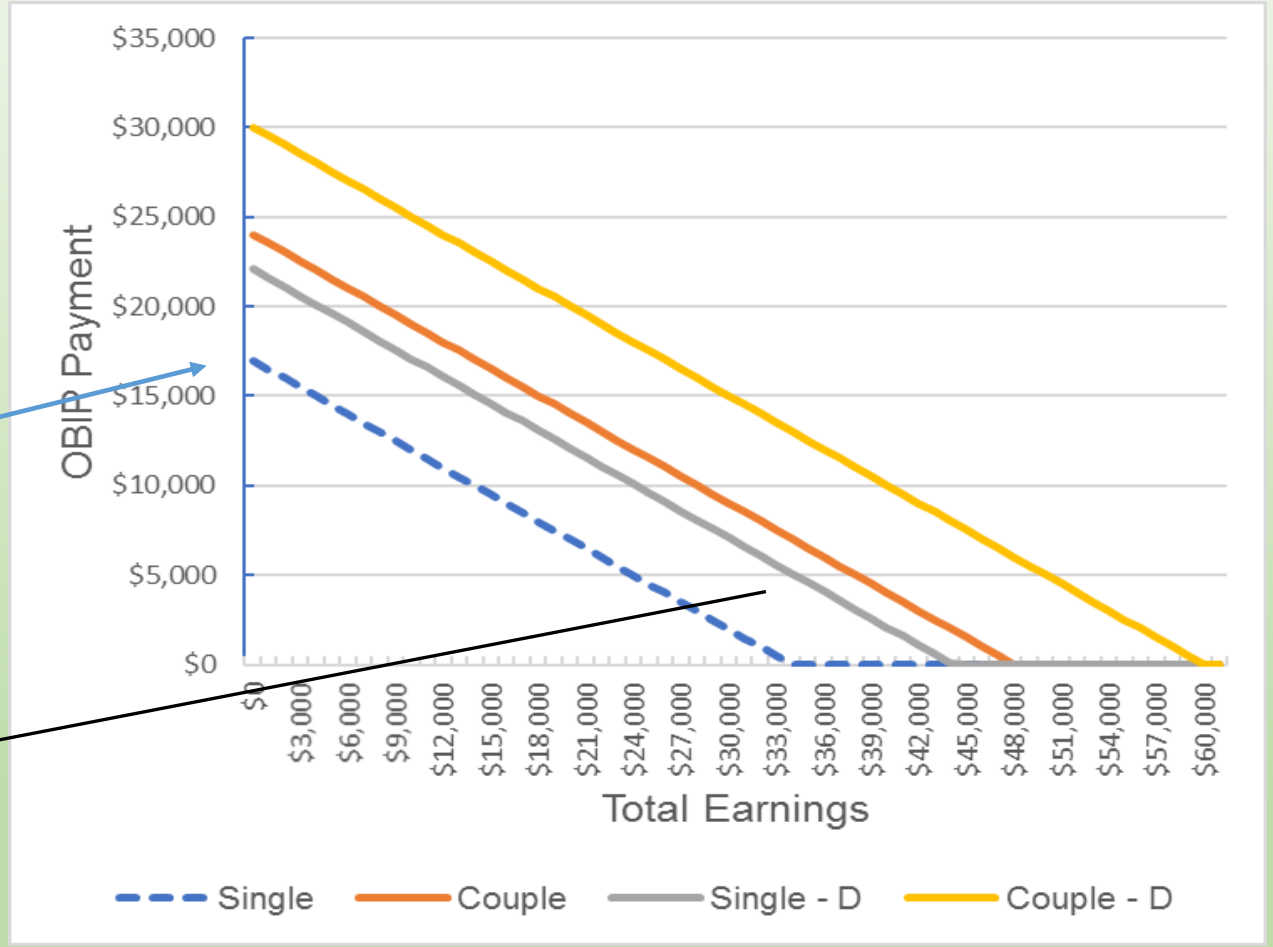
Key lessons for evaluators from Mincome

- Longitudinal survey data can collect a wide array of data generated by designer questions.
- Attrition is will occur and is non-random, implying that statistical correction and control are needed even when the design starts as an RCT.
- Narrow and specific hypotheses support increased experimental control through rigorous sample design.
- Asking questions outside the scope of the experiment and questionnaire, require increasingly artful statistical methods. Confounding will occur as unmeasured external changes affect outcomes.
- Expensive, long term studies try the patience of funders and political sponsors.
- This was the fate of all the major income-maintenance experiments of the seventies.

Ontario Basic Income Experiment (OBIP)

Fast forward to the Ontario Basic Income Pilot (OBIP)

Ontario Basic Income Pilot – Maximum Benefits		
	Single	Couple
Maximum Basic Income Amount	$16,989	$24,027
Maximum Basic Income Amount plus Disability Supplement for one person with a disability	$22,989	$30,027



At $33,000 in earnings, a single person exits OBIP

OBIP is a tax free benefit

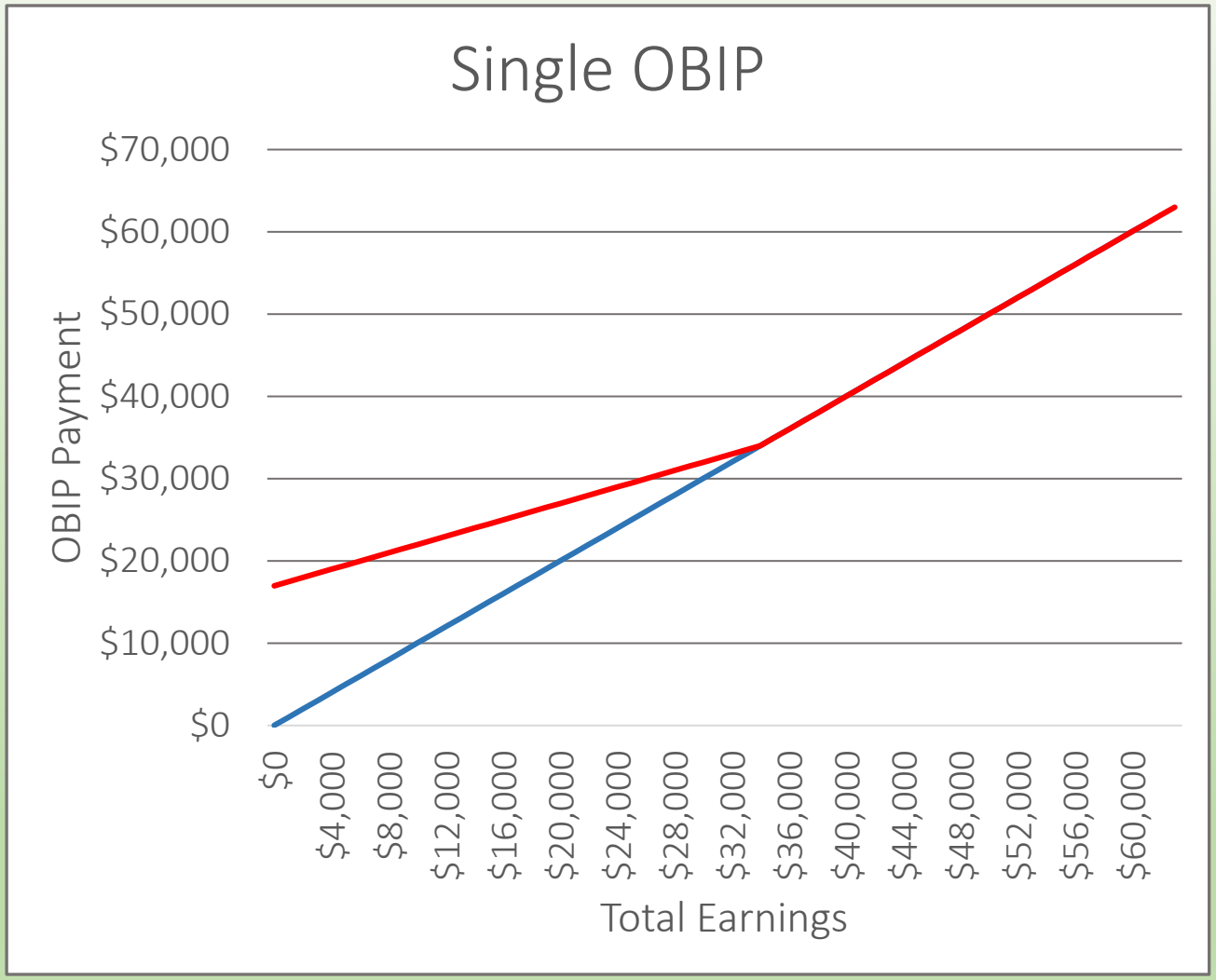
OBIP – Another perspective

For a single person with no income, OBIP provides $16,989 .

As income rises, OBIP payments fall, until income reaches $33,000.

As a negative income tax, every dollar of earned income results in a 50¢ reduction of OBIP.

It always pays to work.



Children were covered under the CCB and OCB

Examples:

- A single individual, without a disability, earning $28,000, will receive and BI payment of be $2989 to make total income to $30,989.
- A couple, with one person disabled, and with part time employment between the two of them generating $16,009, will receive 22,027 in BI with will leave them with a total income of $38,027 .
- A single parent with two children under 6 and no earned income, will receive a BI of $16,989 plus the CCB of $12,800 to reach a total income of $29,789 tax free.
- In the case of the single individual earning $28,000, the combined federal and provincial tax (on earnings) will be about $3,500 reducing their after tax income (with the BI on top) to about $27,500

OBIP vs SA (Ontario Works)

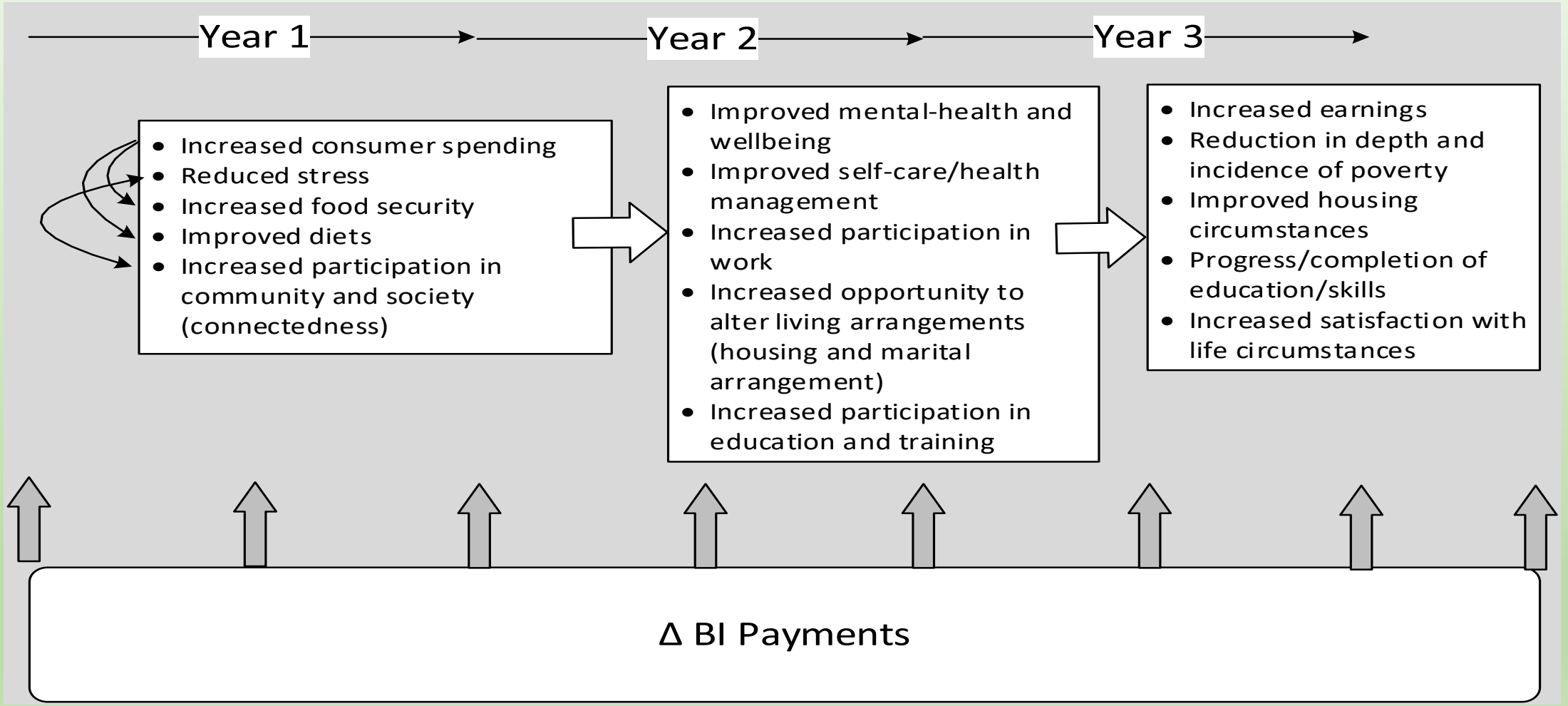
- BI offers more financial support than SA
  - a single person on OW receives $1100/month and has other benefits (health, employment assistance, housing, etc.)
  - OW has a strict asset test (less than $2500), with the principal residence exempted.
  - BI does not evaluate wealth.
- Everyone who files a tax return is eligible for the BI immediately experiences the following:
  - Reduce depth and incidence of poverty (higher income and wider eligibility)
  - Be less intrusive
  - Have an incentive to work

Theory of Change – expected results

Under OBIP, the goals for a BI have broadened, creating a need for complex evaluation designs

- Immediate outcomes (Year 1)
  - Poverty reduction
  - Increased food security
  - Lower anxiety
  - Educational planning and participation
- Intermediate outcomes (Year 2 and 3)
  - Mental health (self-reported and reduced use of services) should show improvement
  - Less contact with the criminal justice system
  - Marital changes, although the nature is hard to predict .
- Longer-term outcomes (at the end of the pilot) should become manifest:
  - Participants will consume better housing
  - Participants will have increased “connectivity” and purchase a wider range of consumer items
  - Educational initial and successful progress
  - Stable work attachment

Logic Model and Theory of Change



Evaluation Design of OBIP

- OBIP was intended to be an RCT, with two main sites (Hamilton and Thunder Bay) in Phase 1, and then with a saturation site (Lindsay) added in the second year (Phase 2).
- An early problem occurred since those on social assistance (SA) have non-financial benefits. Health benefits (HB) were the most important.
- A key tenet of a basic income is that former SA recipients transitioning to the BI need to pay for a range of goods and services formerly covered.
- Everyone moving from SA to OBIP retained their supplementary health benefits.

Design matrix			
Intervention		Comparison	
Single (WO HB)	Couple (WO HB)	Single (WO HB)	Couple (WO HB)
Single (W HB)	Couple (W HB)	Single (W HB)	Couple (W HB)
Single D	Couple D	Single D	Couple D
WO HB – no health benefits, W HB – health benefits on top of OCIP, D – disabled			

Enrolees will participate in period interviews (every 6 months)

A single respondent reports on behalf of the household

The questionnaire intended to require 20 minutes

Continued participation requires sharing tax returns and participation in the follow-up surveys. Comparison group paid for survey participation

University of Manitoba

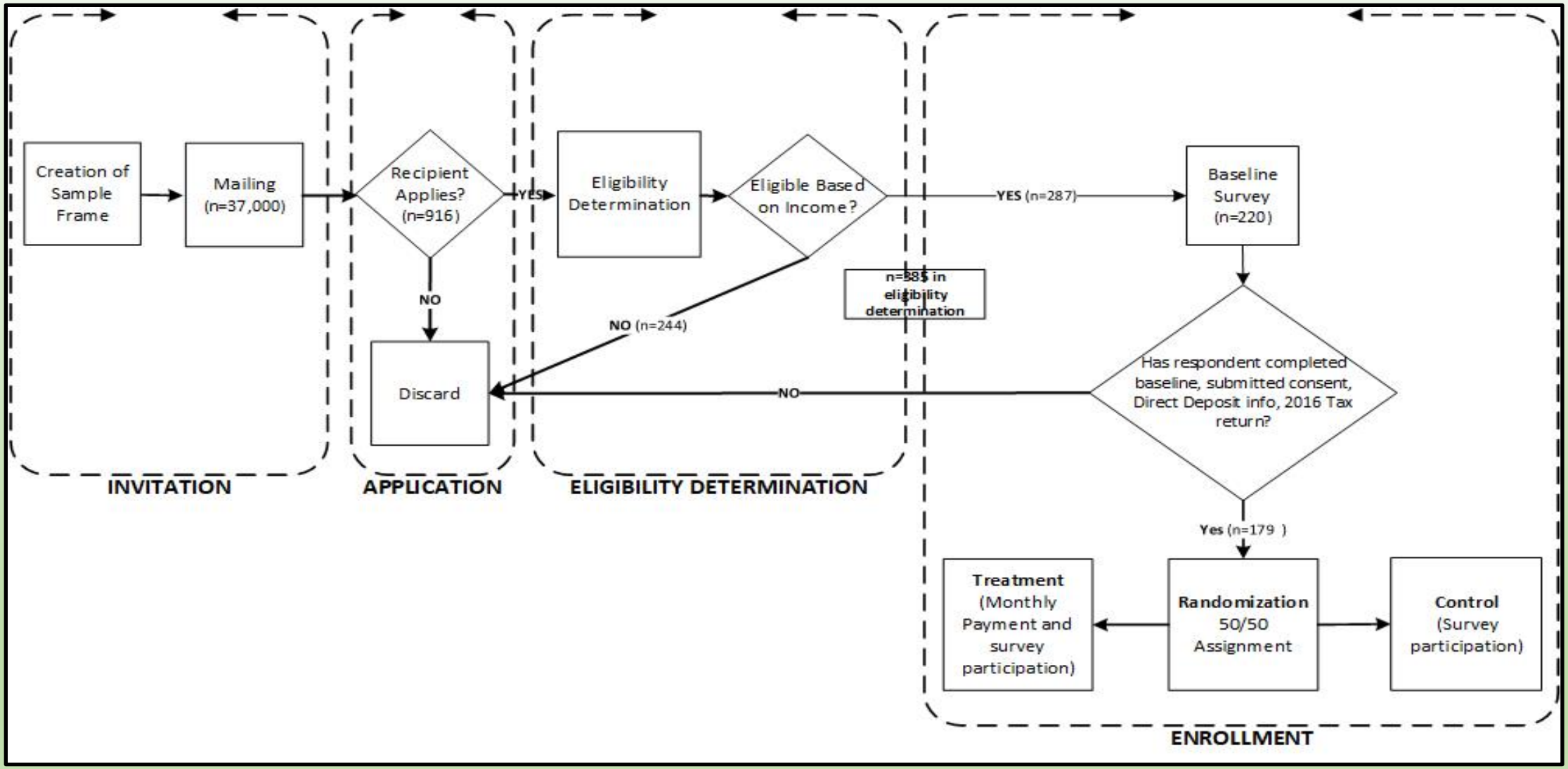
OBIP has two phases

- **Phase 1** (May 2017 – Dec 2017) Baseline
  - Test participation and sample maintenance (Hamilton and Thunder Bay)
  - Assess enrolment procedures
  - Validate questionnaire and survey logistics
  - Guide the development of Phase 2
- **Phase 2** (January 2018 - ?) and included the saturation site – Lindsay
  - Generate periodic samples to support hypothesis testing (Added saturation site at Lindsay)
  - Outcome monitoring
  - Link to health records to assess impacts on mental and physical health
  - Project final cost a province-wide rollout

The realities of sample development

- The target sample for Phase 1 was originally 800+ over the two sites
- Initial advice was to replicate the operation of a basic income, by selecting and inviting eligible respondents from the tax returns. This encountered three problems
  - Canada Revenue Agency opposes the use of tax returns to test policy options
  - The Ministry of Finance, does not have the conduct of the basic income in its mandate
  - Most important, social assistance recipients and many low income households do not file an income tax return.
- The next option was a targeted mailing to low-income areas (census enumeration areas), but this was rejected because this would omit low-income households in high income areas.
- The selected approach was to randomly mail a sample of households in each area, but was very unproductive and costly.
- The final enrolment technique used samples generated by invitations by community organizations.

OBIP Enrolment



Specific Lessons from OBIP

- The enrolment process was mind-bendingly complex and convoluted (Privacy lawyers were determined to earn their pay)
- Invitations to potential respondents read at a second year university level, and never got under grade 10 reading ease.
- The complex theory of change required a marathon questionnaire
- Many concepts used in the questionnaire that seemed clear, created all manner of ambiguity
- The use of sites is an unnecessary complication for evaluating a universal program
- On-line surveys should work in principle, but the fact is that most respondent preferred paper surveys submitted by mail
- Household composition is very hard to record and track
- Single instances can create political liabilities (single gamers, living free in their parent's basements, earn $16,000 to eat cheezies and kill zombies on-line)
- The focus was more on enrolling and not the on-going data collection. Every enrollee creates future obligations to follow-up
- OBIP will become ever more expensive
- The sample distortions and panel irregularities will require complex inferential analysis, negating the analytical value of the

General Lessons Learned from UBI Experiments -

- Both experiments offers important lessons on conducting a basic income pilot
- Politics is paramount
  - With Mincome cost and the lack of results created a political liability and both Manitoba and Canada pulled out.
  - With OBIP, an overwhelming "political need" was start mailing checks fast, with our consequences
    - OBIP became a social program and not a tax program
    - The theory of change reflected a wide range of interests – outcomes are "Hail-Mary"
    - The integrity of the RCT design was thrown overboard as enrolment stalled
    - Casual inference would have needed to rely on "stathocery" (statistical ad-hocery)
- Many SA recipients are "satisfied" with the system. The low income households often mistrust government… this challenges enrolment hard.
- An eventual economy wide roll-out may need to maintain SA as a transition, increasing overall implementation costs.
- Answering the important outcome questions requires a three year (minimum) evaluation period…governments and the public are impatient.

Breakout 2