Challenges and Results in Using Audit Trail data to Monitor Labour Force Survey Data Quality

Justin Francis and Yves Lafortune
Statistics Canada
2016 Methodology Symposium

Outline

- **Labour Force Survey & Audit Trail**

- Objectives

- Results

Labour Force Survey (LFS)

- Monthly household survey (n=56 000)
- Employment, unemployment, related labour information
- Panel survey:
  - 6 panels each in for 6 consecutive months
- Multi-mode collection:
  - In person (CAPI), telephone (CATI) and internet
- Usually 1 person answers for household
 - Proxy responses common (~50%)

Audit Trail

- Detailed paradata on survey collection process with computer-assisted interviewing (CAPI & CATI)
  - Order fields were visited
  - Time stamps and time spent in each field
  - Values entered and edits made

- Very large files with many records for LFS

- Semi-structured in raw form
  - Preliminary work done to convert into structured data

Audit Trail - Cleaning

```
"2012/02/28 8:16:00 PM","Enter Form:1","Key:D00001 100"
"2012/02/28 8:16:00 PM","Metafile name:\\SHEBLA1\Blaise\Production\social\CCHS\CCHS\CCHSComponent.bmi"
"2012/02/28 8:16:00 PM","Metafile timestamp:27 janvier, 2013 15:15:14"
"2012/02/28 8:16:00 PM","WinUserName:klbkbjb"
"2012/02/28 8:16:00 PM","DictionaryVersionInfo:0.0.0.0"
............
:..........
"2012/02/28 8:16:01 PM","Enter Field:OC.OC_Q01","Status:Normal","Value:"
"2012/02/28 8:16:09 PM","Leave Field:OC.OC_Q01","Cause:Next Field","Status:Normal","Value:Ottawa Region"
"2012/02/28 8:16:10 PM","Enter Field:C2.GEN.GEN_Q11B","Status:Normal","Value:"
"2012/02/28 8:16:16 PM","Leave Field:C2.GEN.GEN_Q11B","Cause:Next Field","Status:Normal","Value:819-222-2222"
"2012/02/28 8:16:17 PM","Enter Field:RVC.RVC_N02","Status:Normal","Value:77"
"2012/02/28 8:16:17 PM","Action:Help","Field:RVC.RVC_N02"
"2012/02/28 8:16:20 PM","Leave Field:RVC.RVC_N02","Cause:Move left","Status:Normal","Value:44"
"2012/02/28 8:16:21 PM","Enter Field:RVC.RVC_N02","Status:Normal","Value:"
"2012/02/28 8:16:21 PM","Action:Remark Clicked","Field:RVC.RVC_N02"
"2012/02/28 8:16:40 PM","Action:Remark Changed","Field:RVC.RVC_N02","Remark:Normal visit, no inconsistencies"
"2012/02/28 8:16:40 PM","Leave Field:RVC.RVC_N02","Cause:Next Field","Status:Normal","Value:"
"2012/02/28 8:16:41 PM","Enter Field:Household.RS.RS_Q04","Status:Normal","Value:"
"2012/02/28 8:16:41 PM","Leave Field:Household.RS.RS_Q04","Cause:Next Field","Status:Normal","Value:1"
"2012/02/28 8:16:42 PM","Action:Error Jump","Field:Household.RS.RS_Q04"
............
............

"2012/02/28 8:25:28 PM","Leave Form:1","Key:D00001 100"
```
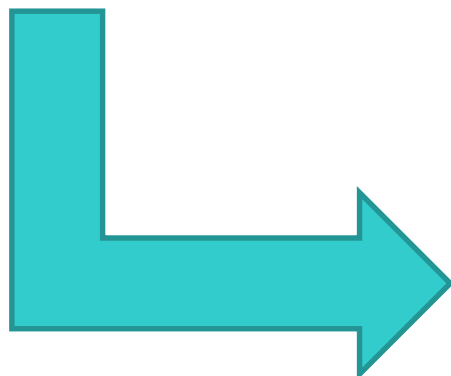

InstrID	VisitSeq	Mode	Month	Field	Duration	VisitType	Entry	Exit
17	1	CATI	201302	N010	1.73	1		2
17	2	CATI	201302	Q100	5.25	1		2
17	3	CATI	201302	Q101	4.88	1		2
17	4	CATI	201302	Q104	1.39	3		
17	5	CATI	201302	Q101	2.5	2	2	1
17	6	CATI	201302	Q102	4.01	1		2
17	7	CATI	201302	Q110	1.84	1		1
17	8	CATI	201302	Q114A	3.94	1		1
17	9	CATI	201302	Q115A	4.05	1		1
17	10	CATI	201302	Q116A	3.44	1		1
17	11	CATI	201302	Q117A	4.71	1		1

Outline

- Labour Force Survey & Audit Trail

- **Objectives**

- Results

Analysis Objectives

- Exploratory analysis

- Gain insights into LFS collection process

- Define quality measures

- Identify possible quality concerns

Defining Quality Measures

- Edit Rates
    - Do some questions require more edits?
    - Does this depend on collection mode?

- Backtracking
    - Where do interviewers reverse directions for a check?
    - Where do they commonly backtrack to fix a specific field?
    - Indicator of unintuitive question flow, confusion or clarification

- Time to complete questions
    - Do some questions often require much longer to answer?
    - Need for interviewer explanation, probing or harder recall
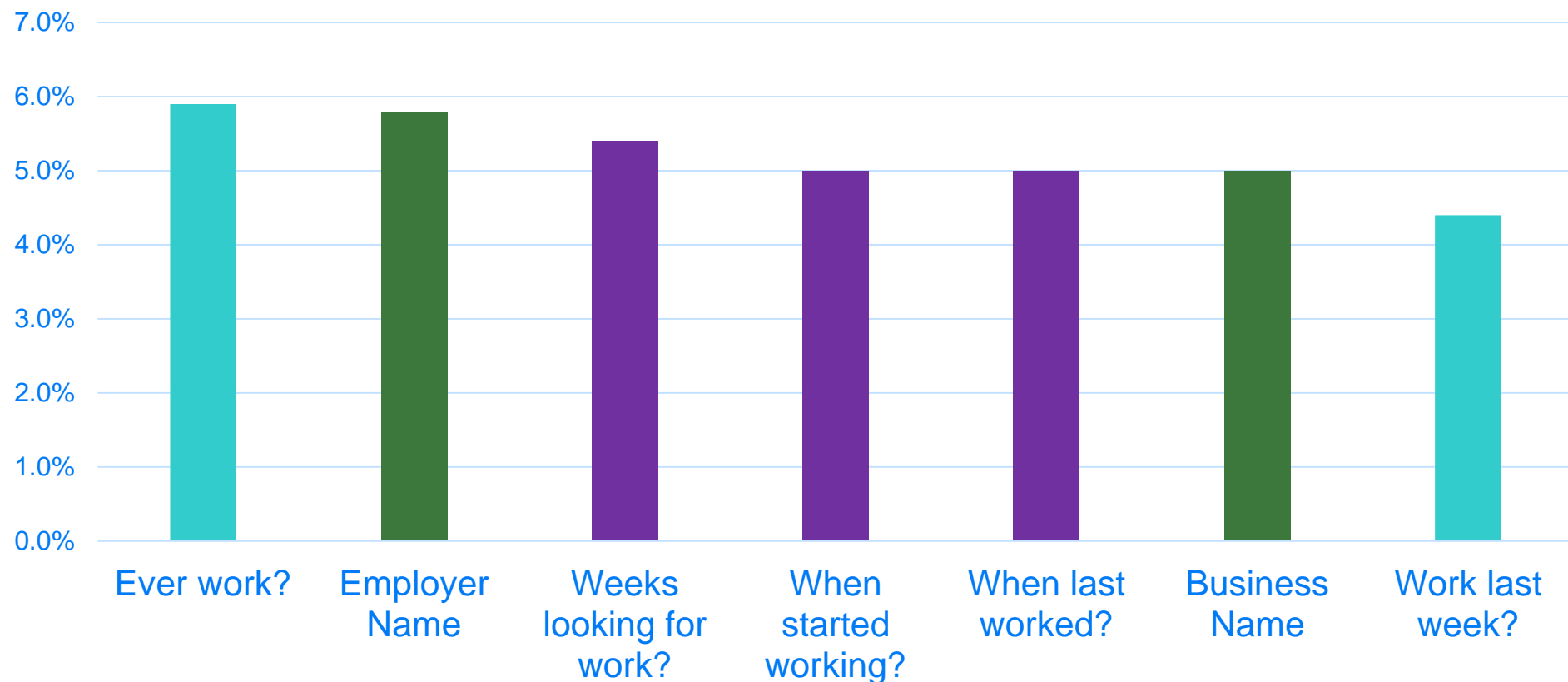
Outline

- Labour Force Survey & Audit Trail

- Objectives

- **Results**

  - Edits

  - Backtracking

  - Blocks with the most backtracking

  - Collection Mode Differences

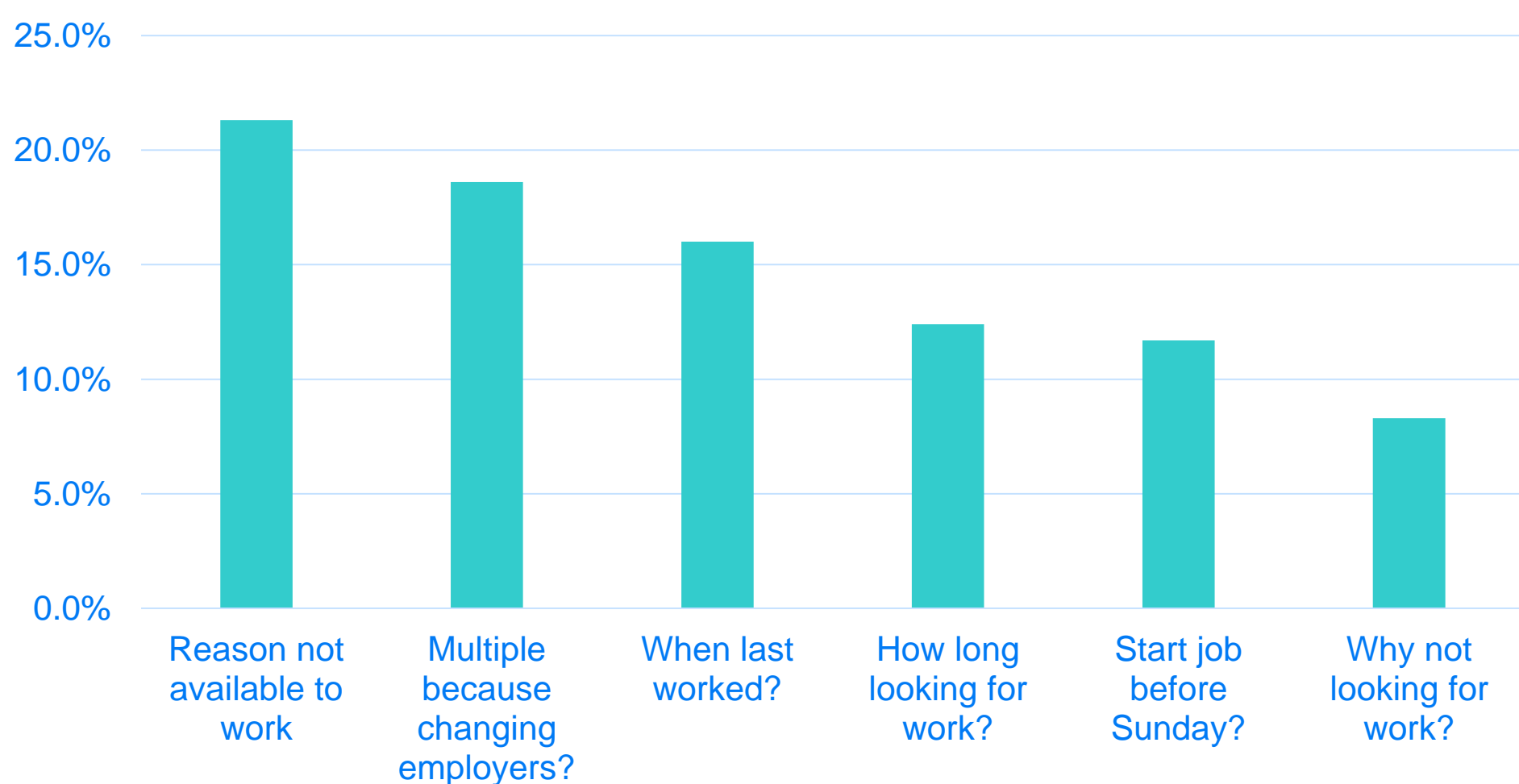 - Time Anomalies

  - Where do we lose respondents?

Edits

- Text fields (typos)
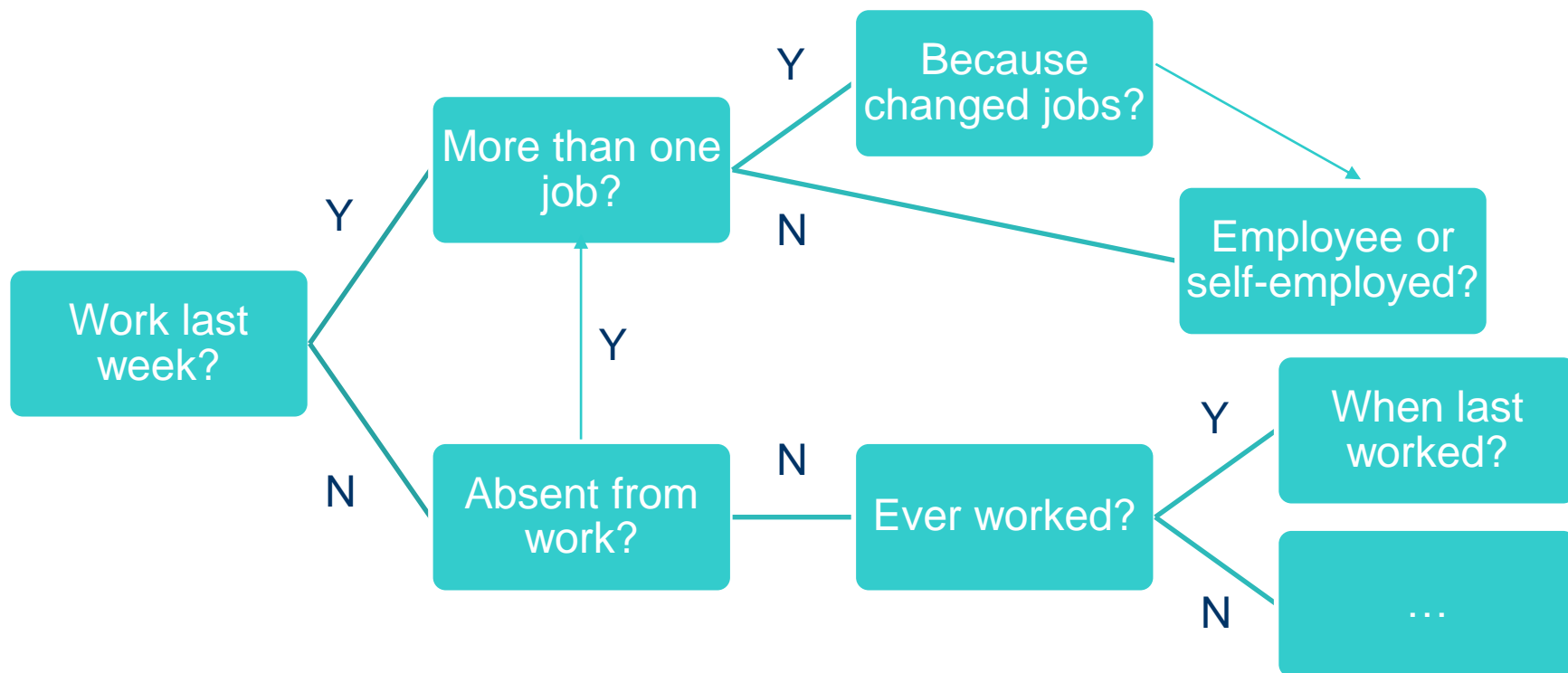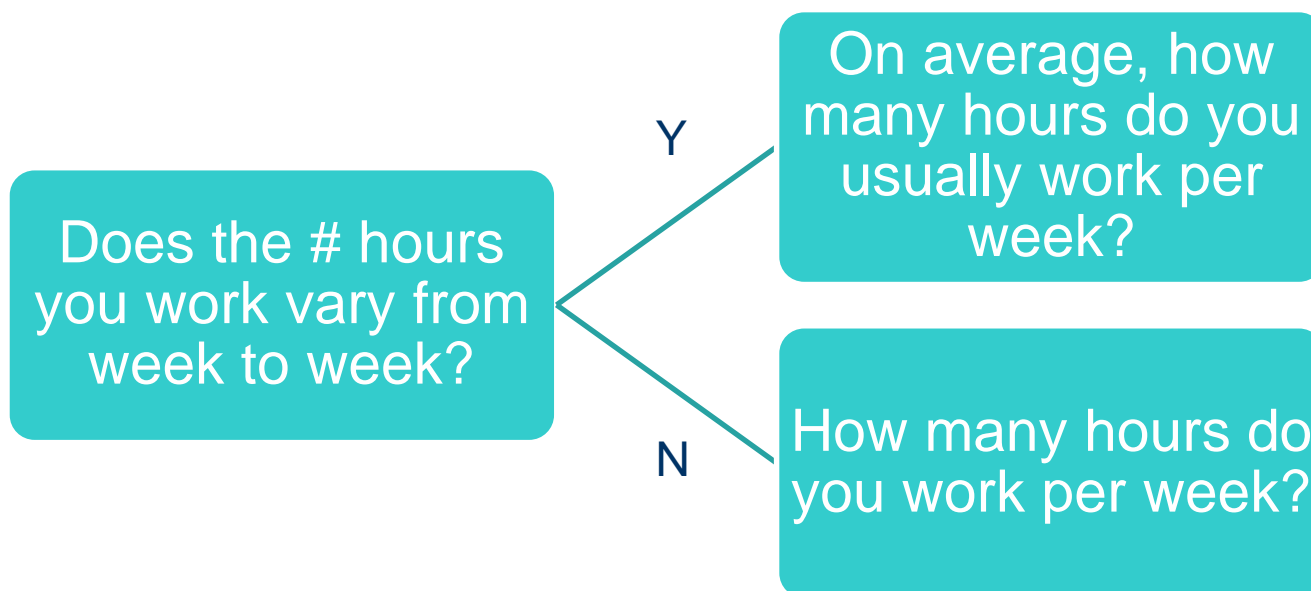- Concept of "working"
- Time Recall

Backtrack changes – Did you work?

Backtrack changes – Hours Worked

Does the # hours you work vary from week to week?

Y → On average, how many hours do you usually work per week?

N → How many hours do you work per week?

Backtrack changes - Wages
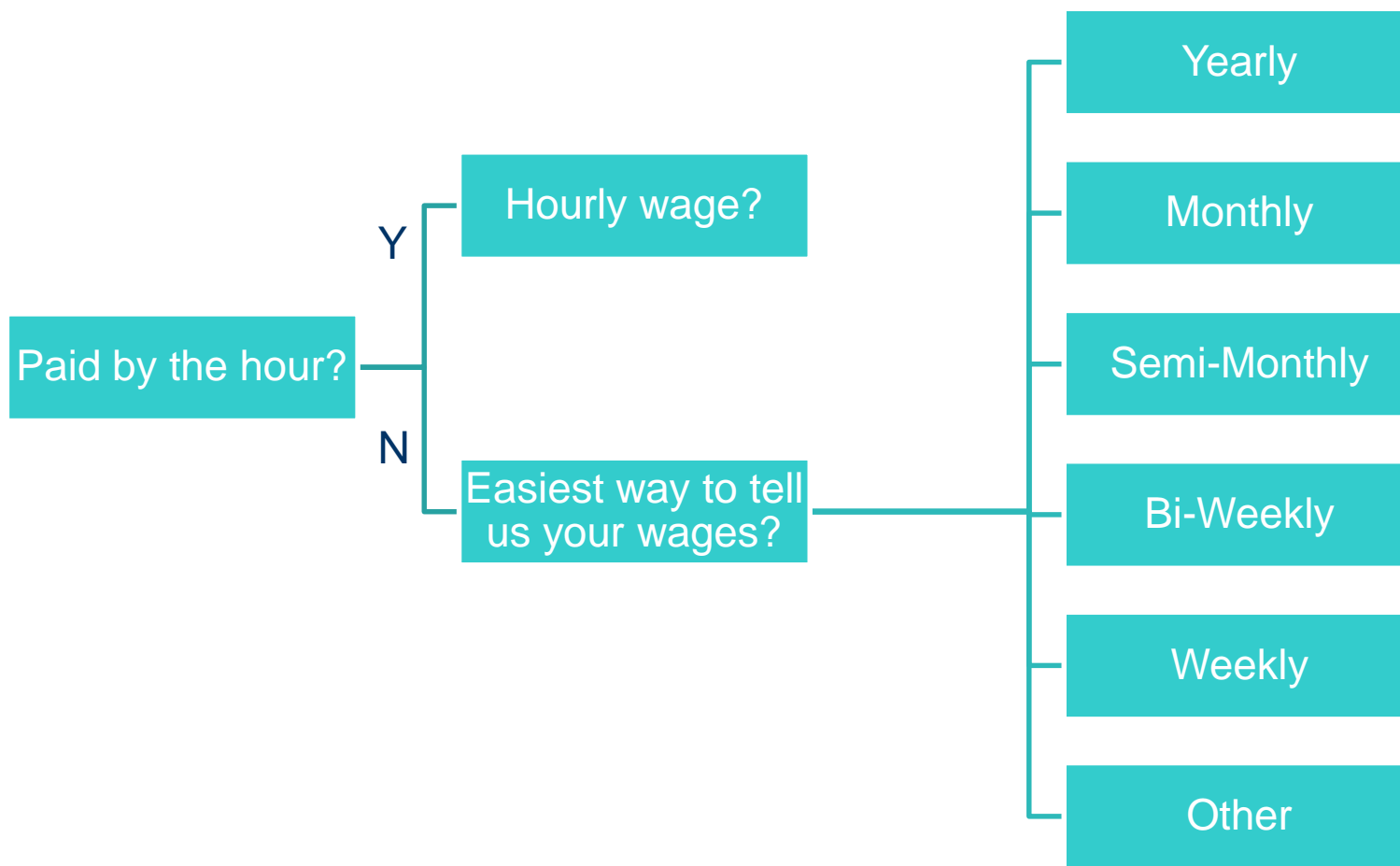
Collection Mode Differences

- Chi-square Test for Independence shows more edits and backtracking in CATI than CAPI
  - Proxy
  - Ever worked?
  - Absent from work last week?
  - Text fields (e.g. business/employer name)

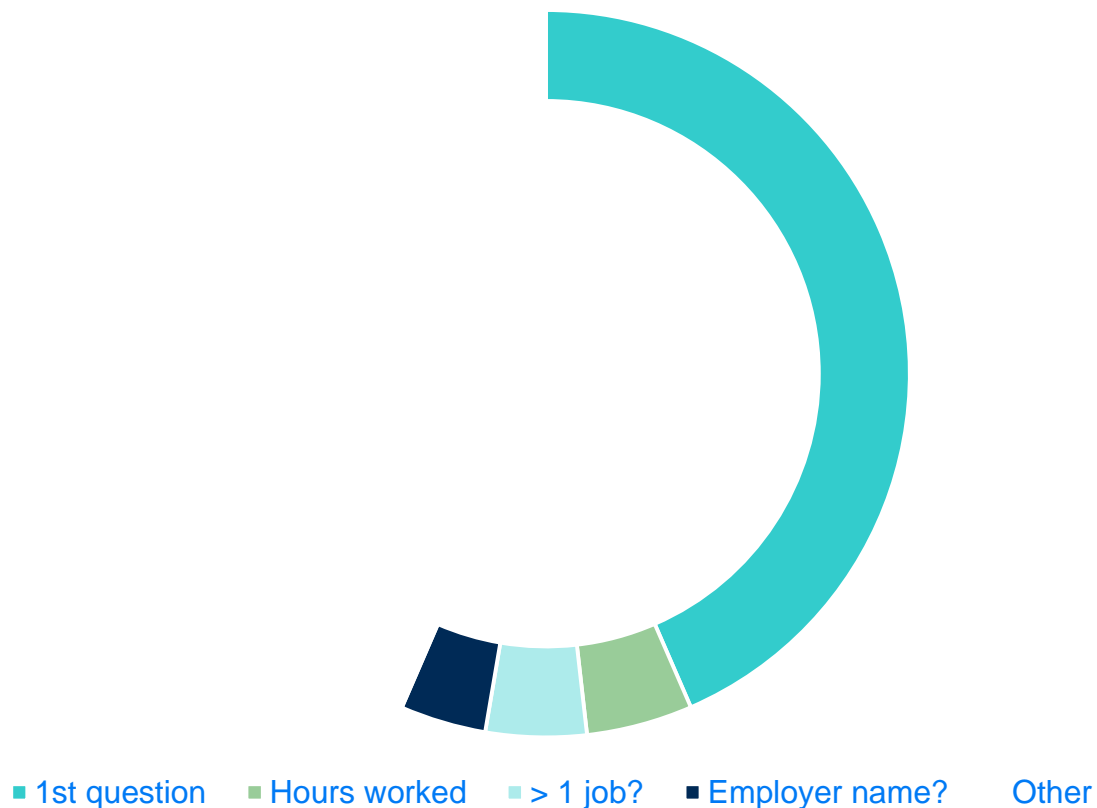- CAPI interviews may be adding value by probing/clarifying

Time Anomalies

- Fields with right-skew and extreme times

- Probing about not working
  - Did you do anything to find work in the last 4 weeks?
  - Could you have worked last week?

- Employment definitions
  - Were you absent from work last week?
  - Are you an employee or self-employed?

- Difficult past recall
  - In what year did you start working?

Where do people drop out?

- Last field visited (excluding edits)

■ 1st question   ■ Hours worked   ■ > 1 job?   ■ Employer name?   Other

Statistics Canada • Statistique Canada

Closing Thoughts

- Audit Trail can provide useful insights

- Some probing questions are helping "fix" responses
  - Respondent or interviewer realizes they are on the wrong path and correct it

- Some question flows and wording may be confusing
  - CAPI interviewers may be adding value by clarifying/simplifying
  - Possible quality or burden issues as collection shifts to CATI & CAWI

- More opportunities for LFS analysis
  - Which types of respondents more likely to need edits in CATI than CAPI
  - Anomalous time data as an indicator of skipped questions

- Audit Trail has value for other household surveys

Thank you / Merci!

- Justin Francis, Statistics Canada

  - Justin.Francis@Canada.ca

- Yves Lafortune, Statistics Canada

  - Yves.Lafortune@Canada.ca