

Development and Programming of an Employment Event History Calendar in the Panel Study of Income Dynamics

April Beaulé, Mary Dascola and Youhong Liu, University of Michigan

1. Introduction

The Panel Study of Income Dynamics (PSID) is a nationally representative longitudinal study of approximately 8700 U.S. families. Since 2003, the PSID has used various Event History Calendars (EHC's) for capturing retrospective reports. The first electronic version of the EHC was built as a stand-alone component to collect employment history of the main respondent and their spouse. In 2007, new childhood health calendar was added. The new calendar was designed to collect information on the most common childhood medical conditions including asthma, diabetes and allergies. This new childhood health calendar was programmed using Visual Basic but integrated into the Blaise application so it appeared seamless to the interviewer. The Blaise DEP (Data Entry Program) called the VB application and data from the childhood calendar was written back to the bdb using a Dynamic Link Library (DLL). Given the success of the childhood health EHC, in 2009 the complex employment Event History Calendar was reprogrammed in a similar integrated fashion.

This paper will discuss the development of the new employment EHC, specifically how we incorporated a number of checks and warnings to the interviewers in order to meet the complex requirements of collecting information about employment and labor force participation.

2. Background

The Panel of Income Dynamics had been on an annual interviewing cycle from 1968 through 1997. Due to budget constraints, the study then went to an every other year interviewing cycle. Interest in the Event History Calendar (EHC) data collection peaked during this transition period, as there was concern about whether respondents would be able to accurately answer questions about moves and especially employment spells over the two calendar years prior to the interviewing year. Building on the on the retrospective memory work of Robert Belli, Ph.D., a cognitive psychologist, the PSID conducted an experiment in 1998 on the use of an EHC to collect data in the core domains used in the PSID including housing moves and employment spells. The experiment showed that the EHC method of interviewing provided more accurate autobiographical retrospective reporting in comparison with the standard question list method.¹

A main focus of the PSID survey is employment, income, housing and wealth. Since spells of employment and unemployment are central to the PSID, that section was re-programmed in 2002 as an EHC instead of a standard question list. When the PSID was reprogrammed in 2002, Blaise offered no method to present a calendar type grid. Therefore, the employment calendar was programmed using Visual Basic with a Microsoft Access database on the back end for data storage. Due to lack of stability issues that we faced with our first attempts using DLL's, the decision was made to keep the stand-alone Access database for storage. The University of Michigan programming staff developed an in-house interface that moved data from Blaise to VB and from VB to Blaise again. The result was that the PSID application was broken into five separate applications and various data sources.²

¹ From PSID website see overview of Calendar Methods Study
<http://psidonline.isr.umich.edu/Data/documentation/ehc/PSIDcalendarMethodsStudy.html>

² Beaulé, A. Leissou E. and Liu. (2007): Experience Using and Event History Calendar in the Panel Study of Income Dynamics, Proceedings of the 11th International Blaise Users Conference, Maryland, United States, September 2007.

3. Prior Experience with EHC Data Collection

The EHC provided the interviewer with a user-friendly interface. Data collection feedback from the field indicated that anecdotally, it did indeed help respondents with recall. The EHC provided the an interviewing advantage of the free flowing format where the both the respondent and the interviewer could move from domain to domain and time period to time period in using a semi-structured format that aided respondent recall. The downside to the unstructured format was a sharp decline in data integrity.

The concept of labor force participation is central to the PSID so that the notions of vacation time, out of the labor force spells, and employment spells must all correspond with each other and all weeks of the year must be accounted for. For example, it is important to know the exact number of weeks worked for a given individual. Elapsed work weeks cover the period from when the employment spell began to when it ended. The number of 'weeks worked' is those elapsed weeks minus 'time away' which can include sick time and vacation time for example. Periods where no employment spells exist are also important and we refer to these periods as time 'out of the labor force' or 'Not Working.' For those times when respondents are out of the labor force, we also distinguish between times when they are looking for employment and when they were not looking for employment.

In the earliest version of the employment calendar, these types of employment or out of the labor force spells were collected but we were unable to provide the hard and soft-consistency checks to the interviewer so they could make on the spot corrections with help from the respondent. The application allowed overlaps between various types of spells and each of those inconsistencies had to be resolved by the data processing staff in order to determine overall weeks worked and time spent out of the labor force.

Each type of spell was collected on a separate tab which made it difficult for interviewers to really see how each of the domains corresponded with each other. There was an attempt to provide the interviewer with an overall timeline at the top of the screen just below the question text, but in order for the interviewer to see the details for any of the spells he or she would have to move to each tab to review those. (Figure 1)

Figure 1: Employment Spells—Early version EHC

 Employment Q-list

BC4. I'd like to know about all of the work for money that you have done for the past two years, from January 1, 2005 to the present. Please include self-employment and any other work that you have done for pay. Start with any job that you had during this time.

BC6. When did you start and when did you stop working for this employer? Please give me all of the start and stop dates if you have worked for (this employer/yourself) more than once.

BC5. What was the name of this employer? [IF NECESSARY: This information will help us to process employment information you gave us. The name itself will never be released as part of data from the study]

[IWER: If no employer name is given, ask for job title or anything that can help identify the job.]

[IWER: Before exiting timelines, probe for any other work for pay, no matter how small.]

R	05	SPR	05	SUM	05	FAL	05	FAL	05	WIN	06	SPR	06	SPR	06	SUM	06	FAL	06	FAL	06	WIN	07	SPR
R	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR

Landmark Events

Residence

Employment Summary

Not Working

Landmark Events Residence Employment Not Working TimeAway

Employment Data Entry Window

R	05	SPR	05	SUM	05	FAL	05	FAL	05	WIN	06	SPR	06	SPR	06	SUM	06	FAL	06	FAL	06	WIN	07	SPR
R	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR

Not Working Summary

Smith Brothers

Parmalat

Craftworks

Job# 4

Starting Time				Ending Time					
Year	Season	Month	3rd of Month	Year	Season	Month	3rd of Month	OK	
Before-DK	SUM	AUG	First	2005	WIN	DEC	Last	Cancel	
Employer	Craftworks							Delete	
Note									

In addition to the internal spell consistency problems, we also faced consistency problems with data collected earlier in the Blaise question list portion of the interview. One of our key variables is current employment status. This variable is a multiple choice question that is asked prior to entry into the calendar. For people who say their employment status is 'working now', they must have current main job in the employment domain in the calendar. Conversely, if the interviewer lists a current main job but listed the person's employment status as retired for example, then either the person's employment status should include 'working now' or the job listed should not be a current job. These types of inconsistencies are best resolved on the spot during the interview.

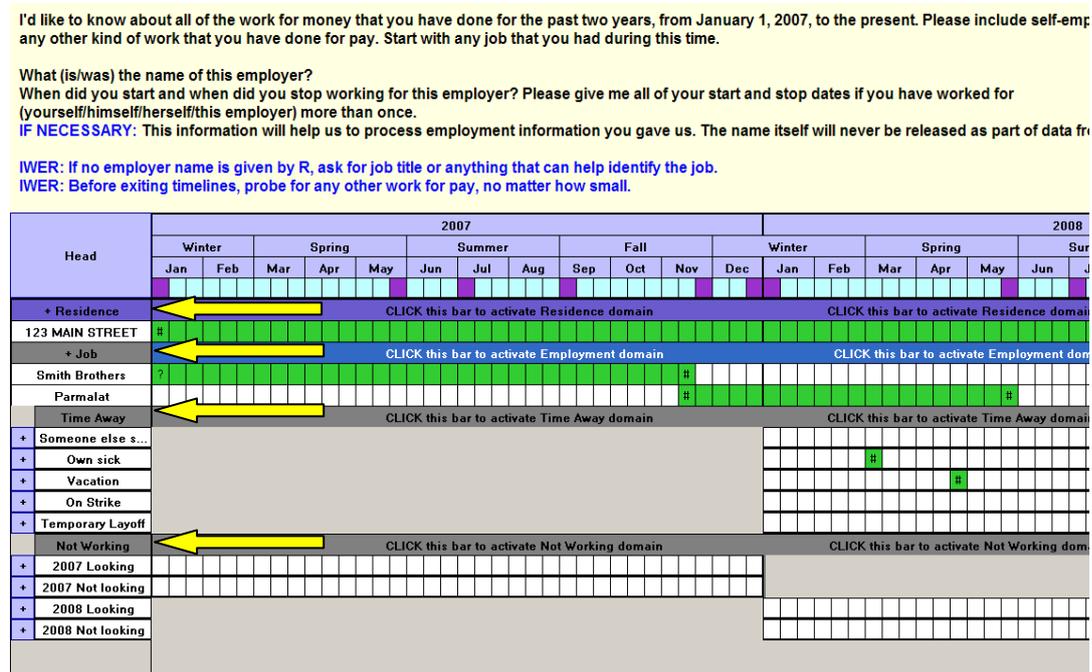
4. Goals for the New Employment EHC

The main goals for of the new version of the employment EHC were to 1) provide a visual guide to the interviewer of the domains and how they closely relate to each other 2) provide the interviewer with checks to help resolve problems during the interview. The new EHC forms were programmed in C#.net.

Some domains need to overlap; spells of 'time away' need to correspond to spells of employment. Some domains should not overlap; spells of 'out of the labor force' should not overlap with spells of employment. The new EHC also needed to provide a way for us to provide consistency checks with the key variable 'employment status' which is located outside the calendar itself. These consistency checks should be easy for the interviewer

to interpret, to respond to, and assist them in soliciting information from the respondent during the interview to resolve data conflicts.

Figure 2: New Version of Employment EHC

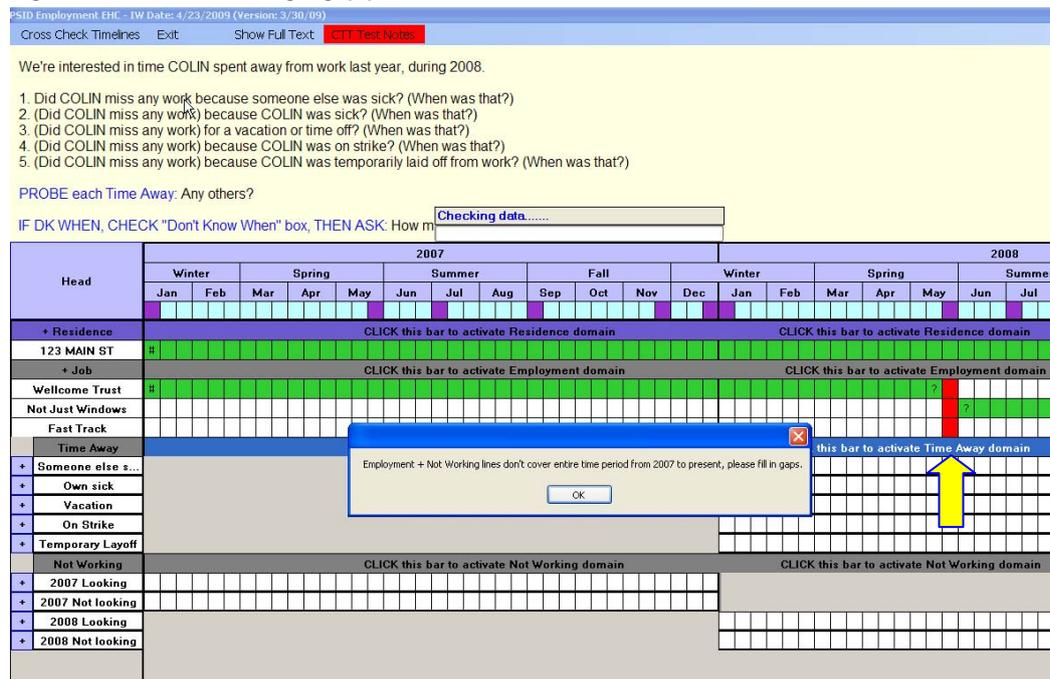


In the newer version of the EHC (Figure 2), you can see that instead of having the domains listed on separate tabs, they are listed all on one screen with the domains listed along the left hand side. This allowed the interviewer to see the rows for each of the domains at the same time and to see where gaps and overlaps are happening.

Since each cell represented 1/3 of a month—we had to allow for some overlap between different types of spells such as employment and unemployment but wanted to warn the interviewer when we had gaps or overlaps occurring. We used several techniques to alert the interviewer in a visual way about the overall integrity of the employment data.

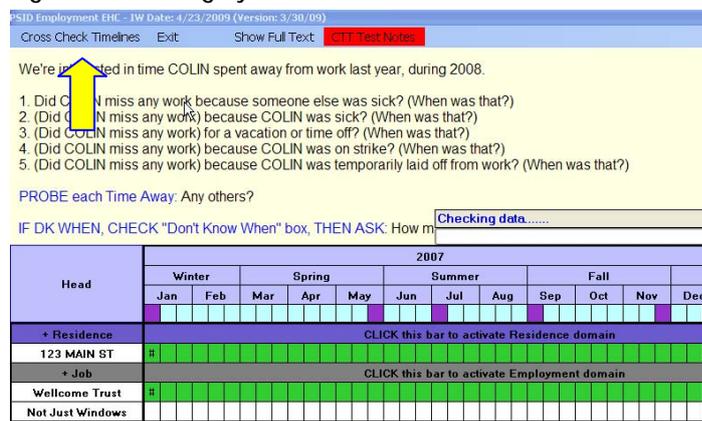
First, we used color to show the interviewer where problems exist. We took our inspiration from traffic signs; green to indicate these portions of the time lines are consistent and red to indicate that a data conflict should be resolved before the interview can move forward. In Figure 3, the interviewer has initiated the data integrity check and the feedback he or she receives is that there are cells, highlighted in red in the last third of May, which are not accounted for. All cells should be accounted for by a period of employment or a period of 'out of the labor force.' In the instrument we have labeled the types of out of the labor force as 'Not Working.' In addition to the problem cells marked in red, the interviewer gets a warning; 'Employment + Not Working lines don't cover the time period from 2007 to present, please fill in gaps.' There are two possible solutions, either the jobs should be concurrent or the interviewer has to mark that time as some type of 'Not Working.'

Figure 3: Checks indicating a gap problem



Since the data entry task in the calendar is semi-structured, we wanted the checks and warnings to be as unobtrusive as possible. The goal was to allow the interviewer free navigation between domains to collect all the information in whatever order without being tripped up by constant real-time rule checking. Our solution was to provide them with a menu option (Figure 4, below) that would initialize the checking once the interviewer felt all the information had been collected and was complete. The interviewer could then make adjustments based on the consistency check feedback and re-check as necessary until all the conflicts were resolved.

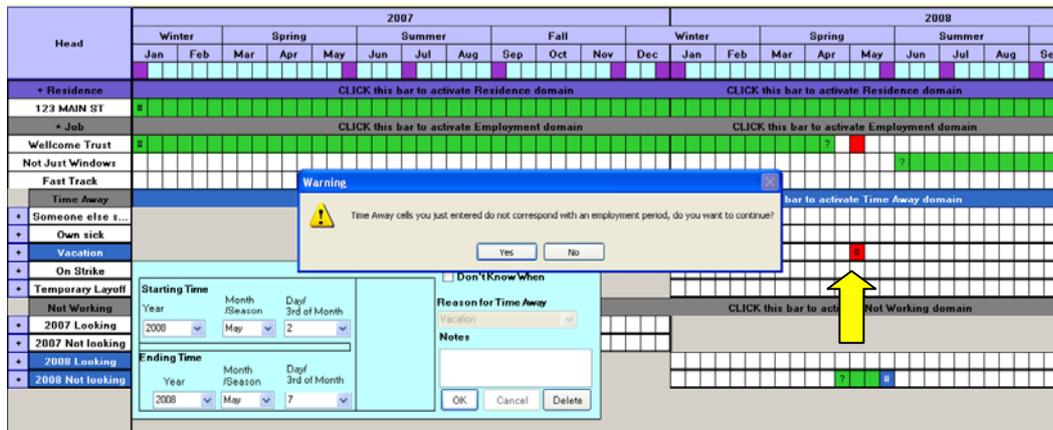
Figure 4: Data Integrity Check Button-'Cross Check Timelines'



For example, a common mistake made in data entry is what we mean by vacation. When respondents think of vacation, they think in terms of the time they went to the beach with the family. When we define vacation, we are talking about times when you have a spell of employment but you are not working because you are on vacation. Respondents will report times of vacation when they are not employed—that time is not considered vacation to us, this is time 'out of the labor force.' In order to help the interviewer understand this difference, we train them about concepts and then we provide consistency checks that remind them that we either have to have a spell of employment to go along with the vacation or not to consider that time vacation but rather, it should be coded as 'out of the labor force' otherwise known as 'Not Working' (Figure 5).

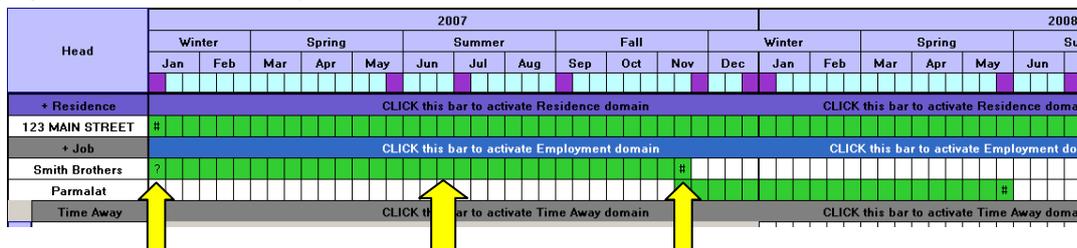
In this scenario, the interviewer has gone to the 'Time Away' Domain and tried to enter vacation for a period of time when the Respondent has no employment cell. The consistency check warns the interviewer 'The Time Away cells you just entered do not correspond with an employment period, do you want to continue?'

Figure 5: Incorrect Reporting of 'Time Away' – Vacation Spell



To further enhance the visual feedback about the timelines, we used symbols at the start and end of each time line and across strings of cells to let the interviewer know whether the timeline had a definite start or end date (actual day, month and year) or whether it was slightly fuzzy (first third of the month), or whether we knew some information but it was less specific. For example, this time line below (Figure 6) shows that the employment spell for 'Smith Brothers' has a fuzzy start date but a definite end date. We used symbols '?' in the first cell to indicate a fuzzy start and '#' to indicate a definite end date. For spells that are much less specific such as two weeks of vacation taken in the year 2008 but the respondent doesn't know when, we fill the cells on that timeline with '?'.

Figure 6: Exact versus Fuzzy Start or End Dates



During PSID post-processing and editing, we spend many hours cleaning up cases where our respondents have overlapping jobs and time away spells. In order to confirm the elapsed 'work weeks', we added a new calculation and routine in 2009. Each cell represents one third of a month. Across the length of the year, each cell represents approximately 1.45 weeks. For respondents with overlapping job spells, we wanted to confirm with them the number of weeks we think they were employed after the employment calendar is completed.

Here is an example of a person with three distinct jobs. Job 1 begins in the second 1/3rd of May and runs through the second 1/3rd of June. Job 2 is from the first 1/3rd of June through the second 3rd of June. The 3rd job runs from the first 1/3rd of June through the last 1/3rd of June. The cells in blue indicate that these cells overlap with cells from another job. The cells in pink indicate that these cells do not overlap with cells from another job.

Figure 7: Calculating and Confirming Elapsed Work Weeks

	May	June	July	TOTWSKJob
--	-----	------	------	-----------

	1	2	3	1	2	3	1	2	3	
Job 1										4*1.45=5.8
Job 2										5*1.45=7.25
Job 3										3*1.45=4.35

Summary Indicators

WTROverlap	NumConcur	NumUnique	Numthirds	Numwks	Rounded
1	3 (Blue)	4 (Pink)	7	7*1.45=10.15	10

Behind the scenes, we process a few items: first we count the number of cells for each job and we multiply them by 1.45 which is the average number of weeks that 1/3rd of a month represents. Then we calculate the maximum number of concurrent cells and the number of unique cells across all jobs and the total is the number of elapsed 1/3rds of months. In this example we have 7 elapsed thirds of months * 1.45 gives us 10.15 weeks, we round that down to 10 weeks and then we confirm this with the respondent.

Q1. : From the information you have given me on your all of your jobs, you worked approximately 10 weeks in 2008, does this sound correct?

The respondent has a chance to let us know if this is a correct reflection of their elapsed ‘work weeks’ and if it is not, they can give us an improved estimate. This confirmation will significantly cut down on hand corrections for ‘work weeks’ calculations during the processing phase.

5. Training Interviewers on the use of an EHC

Interviewer training for the EHC focused on familiarizing interviewers with the new single-screen setup in terms of navigating domains and entering timeline data, and reconciling timeline errors. Data items were entered by choosing a starting cell and then an ending cell. Depending on the item, pop up windows open (Figure 8) so additional information such as employer name can be entered.

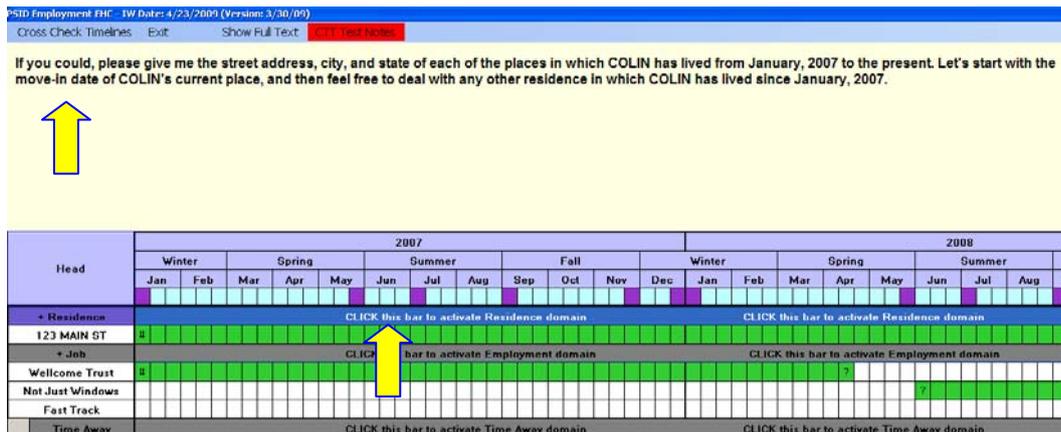
Figure 8: Entering Data Items

The screenshot shows a software interface for entering employment data. On the left is a vertical list of job categories: Wellcome Trust, Not Just Windows, Fast Track, Time Away, Someone else s..., Own sick, Vacation, On Strike, Temporary Layoff, Not Working, 2007 Looking, 2007 Not looking, 2008 Looking, and 2008 Not looking. The main area features a calendar grid for the year 2008, with columns for 'First third', 'Mid third', and 'Last third' of each month. A data entry form is overlaid on the calendar, containing fields for 'Starting Time' (Year: 2007, Month/Season: Jan, Day/3rd of Month: 1) and 'Ending Time' (Year: 2008, Month/Season: Apr, Day/3rd of Month: 1). There is also a 'Current' checkbox and an 'Employer' dropdown menu set to 'Wellcome Trust'. A 'Notes' field is present below the employer dropdown.

Since all domains and timelines are now on one screen, not all information is available for viewing at once. In order to view timelines that ‘fall off’ the screen, interviewers were trained on general navigation techniques such as scrolling and dragging, techniques not needed in the previous EHC. For instance, the two and a half year time period that we are concerned with does not fit entirely on the screen, and interviewers must use horizontal scrolling to get to the latter part of the time period. Depending on the number of timelines, not all can be viewed at one time. In this case, interviewers must use vertical scrolling to get to those timelines towards the bottom of the lower pane. Finally, interviewers were shown how to relocate a blue data entry box by clicking and dragging, in order to better view the box itself or the timeline behind it.

In order to access the question text and timelines for a specific domain, interviewers were trained how to activate a domain by clicking on its labeled ‘domain bar’.

Figure 9 : Using the Domain Bars to Navigate



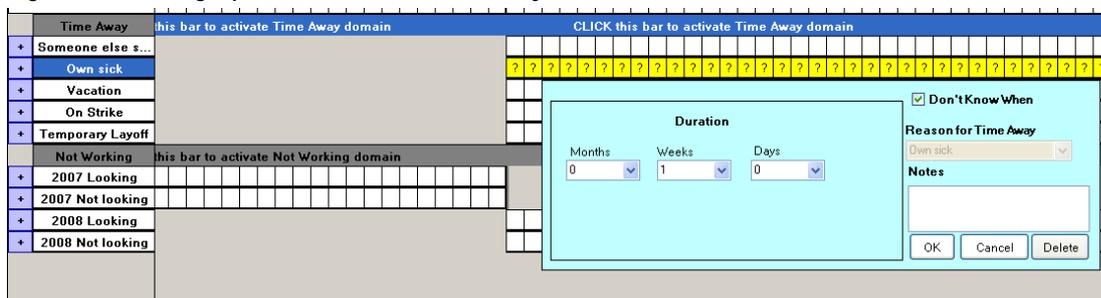
Clicking the domain bar controls presentation of corresponding question text in the top pane and timeline data entry in the bottom pane – neither question text nor data entry for a domain can be accessed without first clicking the domain bar (Figure 9). The interviewer can see which domain is active as the active ‘domain bar’ is highlighted in blue. On the other hand, it is not necessary to activate a domain in order to edit a completed timeline in the domain – that can be done by simply clicking on the title of a timeline regardless of which domain is currently activated.

We relied on a greater amount of hands-on practice during training, in relation to lecture, in order to provide interviewers adequate time to become skilled at moving around the EHC, pointing and clicking to activate domains, and entering timeline data. Additional one-on-one assistance proved to be of great value for many interviewers.

A critical addition to the new EHC is the Cross Check Timelines menu option, which checks for inconsistencies in timelines within and across domains. Inconsistencies are varied and complex. For instance, Time Away timelines are allowed only where there are corresponding Employment timelines otherwise, the Time Away spell is considered invalid. Conversely, Not Working timelines are allowed only where there are **no** Employment timelines. Time Away timelines cannot overlap each other (one can be on vacation or sick, but not both at the same time), nor can Not Working timelines overlap each other (a person is either looking for a job or not looking, but not both at the same time). Timelines may overlap each other by just one cell (third of a month). An exception is an overlap where any start or end date is [DK]. An exception to the exception is a Time Away cell that falls within a period of No Employment, which is never valid, even if a start or end date is [DK].

Below is an example (Figure 10) of entering a period of sick time for the previous year. The respondent knows they had a week of sick time in 2008 but the do not know when. The timeline for this row is marked in yellow with ‘?’’s. In the database, we know that we have a period of own sick time in 2008 and the duration is one week.

Figure 10: Entering a period of ‘Sick Time’ for last year ‘Don’t Know When’



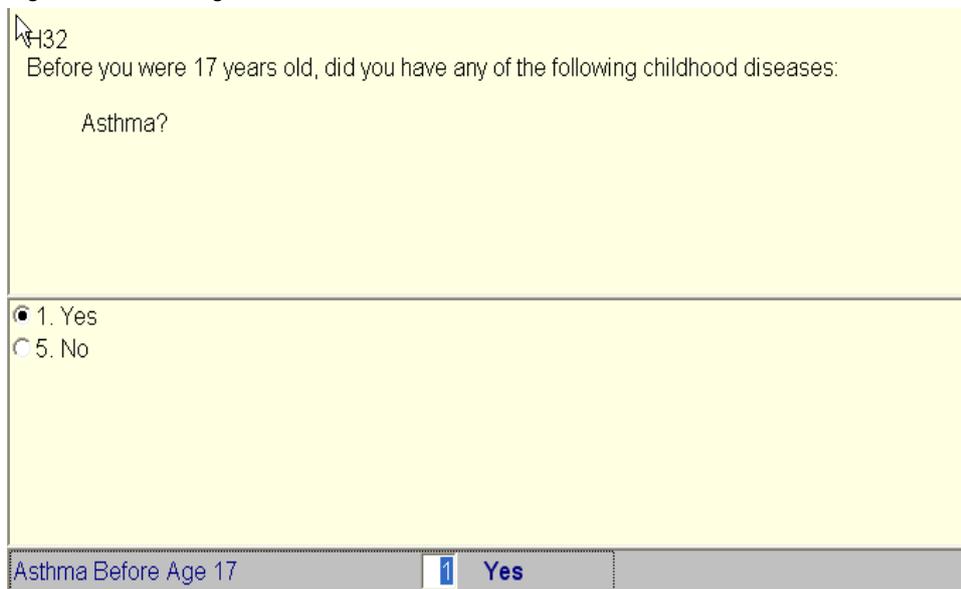
Because of the number and complexity of errors, training emphasized the use of Cross Check Timelines, rather than memorization of all possible inconsistencies and what order the program checked for these. Interviewers were trained to use Cross Check Timelines frequently during the process of moving through the domains, rather than waiting until all domains were completed and all timelines entered. It was recommended to fix an error, by

verifying with the respondent, as soon as the error was presented, and then to run Cross Check Timelines immediately to make sure that the error was indeed resolved.

6. Experimenting with EHC using Dynamic Link Libraries

In 2007 we incorporated the childhood health calendar into the application. The questions focused on childhood health conditions. Since the reference period focused on the respondent's early life, the calendar method of data collection was chosen. Instead of having a secondary data storage file, a DLL call was used that passed control to the subroutine (Visual Basic calendar) if the respondent confirmed they had any of the childhood health conditions listed. When conditions were endorsed in the screening section (Figure 11) then control would pass from Blaise to the EHC where follow up questions about when the respondent had the condition were asked. If no conditions were endorsed by the respondent, Blaise would skip the calendar subroutine and continue on to the next field on route.

Figure 11: Screening Question in Blaise for the EHC Childhood Health Calendar



The screenshot shows a Blaise interface with a yellow background. At the top left, there is a mouse cursor and the number '132'. The main text reads: 'Before you were 17 years old, did you have any of the following childhood diseases:'. Below this, the word 'Asthma?' is displayed. At the bottom of the question area, there are two radio button options: '1. Yes' (which is selected) and '5. No'. At the very bottom of the screenshot, there is a grey bar containing the text 'Asthma Before Age 17' on the left, a small icon in the center, and the word 'Yes' on the right.

The data collected in the calendar was then sent back to the Blaise bdb via the DLL. The use of the DLL for handling data movement and storage was extremely successful. We had no cases with missing or anomalous calendar data.

7. Programming Logistics

The employment EHC DLL is written in C# .net 2.0. In addition, BCP 4.8 is utilized. The first step in programming the EHC is to create data structures in Blaise that work with all the domains in the external application.

- Bemploy (Figure 12) defines employment time line data. It stores start and end date, employer name and a few other pieces related information. The field xXML stores all other data in the block for quick data saving and retrieving in a C# form. xDisplay and sSpanDisplay are constructed in the C# program for later text displays in Blaise fields. Array [1..10] of Bemploy indicates that we are allowing a maximum of 10 employment time lines
- BtimeAway(Figure 13) has the similar data structure as Bemploy. Five arrays are defined to represent 5 time away reasons: Someone Else Sick, Own sick, Vacation, On Strike and Temporary Layoff.
- BnotWork behaves similar as Btimeaway.

Figure 12: Employment Blaise Block Definition

```
BLOCK BEmploy
  Fields
    AQSN : TAQSN
    sXML /"Xml string for easy alien router programming":string[2000]
    sDisplay /"display for summery screen" : string[1000] {part of xXML}
    sSpanDisplay /"display for summery screen" : string[1000] {part of xXML}
    sYear /"Start Year " : 1901..9997,DK,EMPTY
    eYear /"End Year " : 1901..9997,DK,EMPTY
    sMonthSeason /"Start Month/Season (start month)" : 1..24,DK,EMPTY
    eMonthSeason /"End Month/Season (End month)" : 1..24,DK,EMPTY

    sDay / "Start Day (start day of the event)": 1..43,DK,EMPTY
    eDay / "End Day (End day of the event)": 1..43,DK,EMPTY
    Employer / "Event Description": string[100],DK,EMPTY

    Notes / "Event Notes": string[200],DK,EMPTY
ENDBLOCK

EmploySection : Array [1..10] OF BEmploy
```

Figure 13: Time Away Blaise Block Definition

```
BLOCK BTimeAway
  Fields
    AQSN : TAQSN
    sXML /"Xml string for easy alien router programming":string[2000]
    sDisplay /"display for summery screen" : string[1000] {part of xXML}
    sYear /"Start Year 1901..9997,DK,EMPTY
    eYear /"End Year " : 1901..9997,DK,EMPTY
    sMonthSeason /"Start Month/Season" : 1..24,DK,EMPTY
    eMonthSeason /"End Month/" : 1..24,DK,EMPTY
    sDay / "Start Day":1..43,DK,EMPTY
    eDay / "End Day ":1..43,DK,EMPTY
    Reason / "Reason for Time Away": 1..5
    DKWhen : 0..1
    DKWhenMonths : 0..12
    DKWhenWeeks : 0..3
    DKWhenDays : 0..31
    Notes / "Event Notes": string[200],DK,EMPTY
ENDBLOCK

TimeAwaySection1 : Array[1..5] of BTimeAway
TimeAwaySection2 : Array[1..5] of BTimeAway
TimeAwaySection3 : Array[1..5] of BTimeAway
TimeAwaySection4 : Array[1..5] of BTimeAway
TimeAwaySection5 : Array[1..5] of BTimeAway
```

The second step is to create a C# form to work with the Blaise data structure.

- Data grid view controls are utilized to represent domain cells
- Panels with different controls are used for data entry
- Other month and date user controls are utilized for easy month/day selection
- .net dataset is deployed for easy data storing and retrieving ; the xXML field is read into the dataset

Figure 14: C# Form Design View

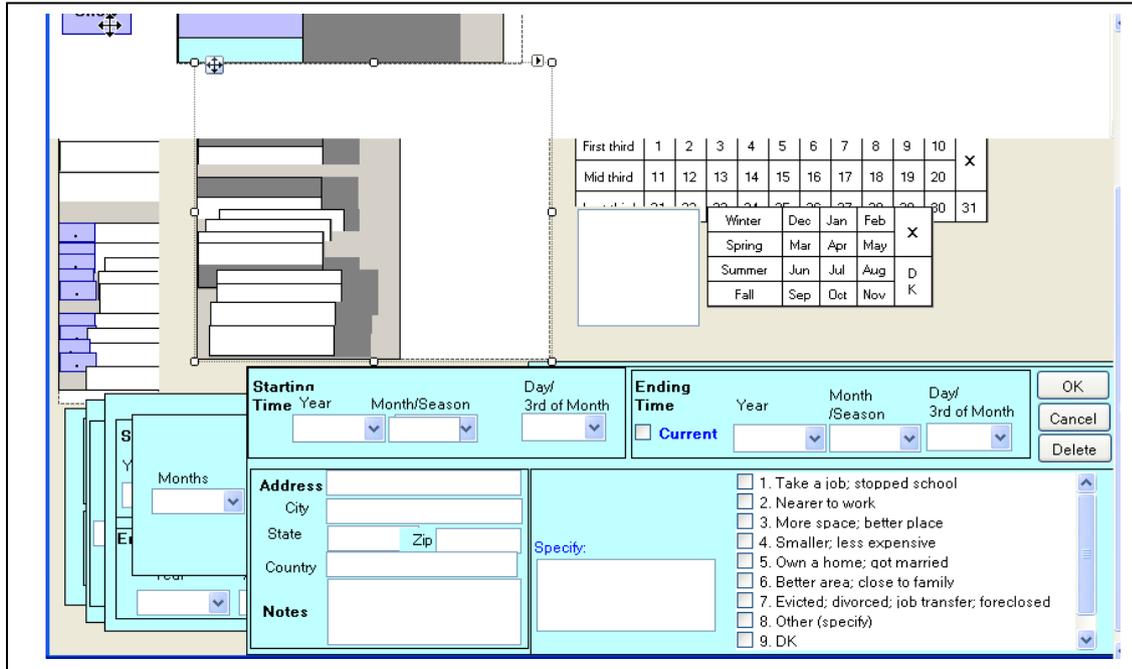
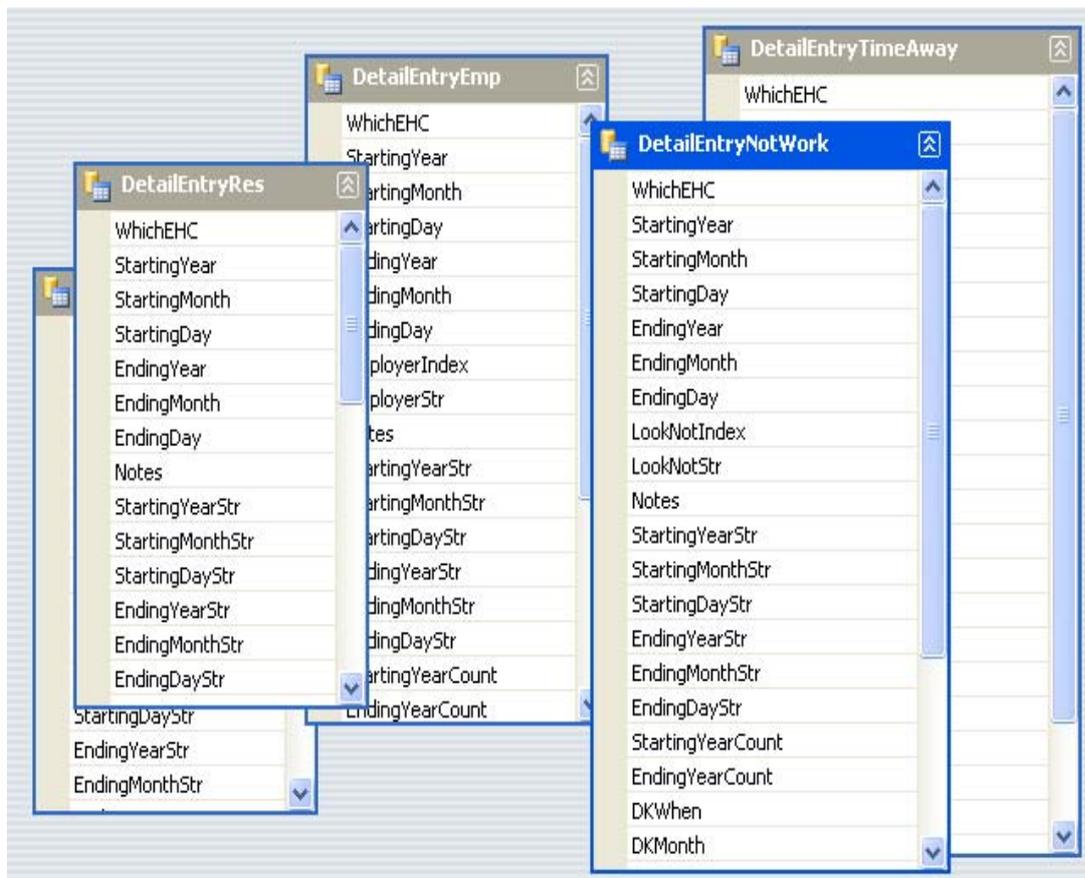


Figure 15: Dataset for all Domains



The final and critical step is to save and check the EHC data for consistency. Once the interviewer presses 'Cross Check Timelines', the program will save data to Blaise bdb. At the same time, it performs various consistency checks across time lines in all domains.

For example, NotWorkValid is used to check for overlapping periods between NotWork and Employment lines. The .net program will set NotWorkValid to '1' if it fails the check and then passes the value to Blaise. In Blaise, the following consistency check (Figure 16) is triggered upon exiting EHC.

Figure 16: Check for overlapping Timelines

CHECK NotWorkNotValid = 0
InVOLVING (BCDE1, EmployEHCRouter.EHCINTRO)
 "One or more Employment and Not Working lines overlap for more than one cell, please reconcile and correct."

Upon exiting EHC, the DLL sets the entry field value in Blaise to '2', so it prevents EHC from opening automatically when interviewer lands to the field next time. If interview wants to edit the EHC, he/she will need to change the value to '1'.

In addition to saving data to Blaise blocks, the data is also saved to an xml file. The xml is essential because the Blaise audit trail function does not capture any data collected in an external application.

8. Conclusion and Summary

Event History Calendar data collection may be superior method over the standard question list method for certain types of question series. The challenge for the PSID project has been learning by trial and error how to

take advantage of this collection method without compromising data quality and integrity. Based on our past experience developing an early version of an EHC for employment, we made a number of improvements including visual clues, consistency checking and clear warnings to the interviewer about potential problems. We have also incorporated 'on-the-fly' verification checks with the respondent so that we have a complete and clearer picture of their employment history.

Data conflicts can be resolved with the help of the respondent during the data collection process. This will mean more accurate coding of weeks of employment and out of the labor force spells and will reduce the numbers of hours spent post-processing trying to resolve common consistency errors.

9. References

From PSID website see overview of Calendar Methods Study. Retrieved May 31st 2007 from World Wide Web:

<http://psidonline.isr.umich.edu/Data/documentation/ehc/PSIDcalendarMethodsStudy.html>

Beaule, A. Leissou E. and Liu. (2007): Experience Using and Event History Calendar in the Panel Study of Income Dynamics, Proceedings of the 11th International Blaise Users Conference, Maryland, United States, September 2007.

Belli, Robert F. and Stafford, Frank P., Calendar and Time Diary Methods: Measuring Well-Being in Life Course Research, Sage Publications 2008-09.

Belli, Robert F., Stafford, Frank P. and Shay, William P. "Event History Calendar and Question List Survey Interviewing Methods: A Direct Comparison." Public Opinion Quarterly, Vol. 65, p. 45-74, 2001.

Hagerman, J. and Kannan, H. (2003): The "Multiple Application Interface" with Blaise and Visual Basic, Proceedings of the 8th International Blaise Users Conference, Copenhagen, Denmark, May 2003.