IHIS: Facilitating use and analysis of NHIS data, 1969-present

Miriam L. King

Andrew C. Ward

Pamela Jo Johnson

ihis

Outline

- Overview

- Background to IHIS

- IHIS Collaborators

- Key components to integration project

  - Harmonization

  - Documentation

  - Dissemination

- IHIS data releases

Overview

- The Integrated Health Interview Series (IHIS) is a five-year NICHD-funded collaboration at U of MN
  - School of Public Health
  - Minnesota Population Center

- Harmonize, document, and disseminate publicly-available data from NHIS for the period from 1969 to the present

Utilization of NHIS Data

- Empirical studies published 1978 – 2003 (n = 999)
  - □ Search "National Health Interview Survey"

- Random sample of 100 studies using NHIS

- Reviewed studies and tabulated the data files used
  - □ 70% used a single year of data
  - □ < 25% used data from before 1986
  - □ No studies used data from all available years

Challenges to NHIS Data Use

- **Complexity discourages use of multiple years of data**
- **Multiple codebooks and files**
  - Variable locations, coding schemes, and response categories may change over time
- **Critical challenges include:**
  - Managing a large number of distinct files
  - Making key documentation easily accessible for each survey year
  - Harmonizing various file structures, sample weighting schemes, and variables

IHIS Inventory (unintegrated)

- **Current emphasis is 1969 - 2006**
  - ☐ Data files: >400
    - ■ Range = 5 to 16 files per year
    - ■ Levels: household, family, person, episodes

  - ☐ Variables: ~34,000 unique variable names
  - ☐ Household records: ~1.5 million
  - ☐ Person records: ~4 million
  - ☐ Lines of data: ~ 20 million
  - ☐ Documentation pages: Too many to count...
    - ■ In 2000, for example, there are over 3,000 pages of documentation describing 2,451 unique variables

Integrating the NHIS

- IHIS builds on model of the Integrated Public Use Microdata Series (IPUMS)
  - see www.ipums.org
- IPUMS spurred new research using harmonized microdata of U.S. census from 1850-2000
- IHIS enhances the value of NHIS data and promotes new research
  - Access to harmonized data and comprehensive documentation from entire period
  - Consistent comparisons across 40 years of dramatic change in public health

IHIS Components

- **Harmonization**

- **Documentation**

- **Dissemination**
  - http://www.ihis.us

Harmonization: Translation Table

	A	B	C	F	S	AH
1	code	label	syntax	ih1969	ih1982	ih1997
2	rectype	MARSTAT		P	P	P
3	columns	Legal marital status		60	88	78
4	norecode			0	0	0
5	hide					
6	proj1			P60	P48	P40
7	proj2					
8	proj3					
9	<tt>					
10	0	NIU	NIU		0=Under 14 years	0=Under 14 years
11	0			0=Under 17 years		
12	1	Married	Married	1=Married		
13	2	Married - Spouse present	Married - Spouse present		1=Married - Spouse in household	1=Married, Spouse in household
14	3	Married - Spouse not in household	Married - Spouse not in household		2=Married - Spouse not in household	2=Married, Spouse not in household
15	4	Married - Spouse in household unknown	Married - Spouse in household unknown			3=Married, Spouse in household unknown
16	5	Widowed	Widowed	2=Widowed	3=Widowed	4=Widowed
17	6	Divorced	Divorced	4=Divorced	4=Divorced	5=Divorced
18	7	Separated	Separated	5=Separated	5=Separated	6=Separated
19	8	Never married	Never married	3=Never married	6=Never married	7=Never married
20	9	Unknown marital status	Unknown marital status		7=Unknown	9=Unknown marital status
21	</tt>					
22	<univ>					
23	ih1969=1981	1969-1981	Persons age 17+.			
24	ih1982=2003	1982-2003	Persons age 14+.			
25	</univ>					
26	<Notes>					
27	Comments					

Data Dissemination

Variable Availability

Variable Availability in IHIS

General Health Variables (P) [top]

Variable Name	Variable Label	Codes	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	78	77	76	75	74	73	72	71	70
BDDAY2WK	Bed disability days in past 2 weeks	codes	.	.	.	.	.	.	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
BEDAYR	Bed disability days, past 12 months	codes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.
BEDAYR2	Bed disability days in past 12 months, intervalled	codes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.
BMI	Body mass index	codes	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BMICALC	Body Mass Index, calculated from publicly released height and weight variables	codes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.
BWGTGRAM	Birth weight, in grams	codes	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BWGTLB	Birth weight, in pounds	codes	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BWGTOZ	Birth weight, ounces only	codes	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
DESIREWGT	Desirable body weight	codes	.	.	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HEALTH	Health status	codes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
HEIGHT	Height in inches without shoes	codes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.
HEIGHTM	Height in meters	codes	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	.	.	.	.	.	.	.	.	.
HSTATYR	Health status compared to 1 year ago	codes	X	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
RADAY2WK	Other restricted activity days in past 2 weeks	codes	.	.	.	.	.	.	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
RADAY2WKRSID	Other restricted activity days in past 2 weeks - Residual Measure	codes	.	.	.	.	.	.	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	.	.	.	.	X	X	X	X	X	X
RADAY2WKTOT	Total restricted activity days in past 2 weeks	codes	.	.	.	.	.	.	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.
SLDAY2WK	School loss days in past two weeks	codes	.	.	.	.	.	.	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
SLDAYR	School loss days, past 12 months	codes	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
WEIGHT	Weight in pounds without clothes or shoes	codes	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.
WEIGHTK	Weight in kilograms	codes	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	X	.	.	.	.	.	.	.	.	.
WLDAY2WK	Work loss days in past 2 weeks	codes	.	.	.	.	.	.	.	.	.	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
WLDAYR	Work loss days, past 12 months	codes	X	X	X	X	X	X	X	X	X	X	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Medical Care Variables (P) [top]

Variable Description & Comparability

INTEGRATED HEALTH INTERVIEW SERIES

HOME | DATA | DOCUMENTATION | VARIABLES | CONTACT

Variable Description

BEDAYR2

Bed disability days in past 12 months, intervalled

Availability
1976-2005

Universe
1976-1996: All persons.
1997-2005: Sample adults age 18+.

Codes

Description
BEDAYR2 reports, in broad intervals, the number of days during the preceding 12 months that illness or injury kept the person in bed for more than half the day, including days while hospitalized.

Comparability — General
For 1976-1981 versus 1982-2003, there were subtle differences in the initial question used to elicit the information for BEDAYR2. For the earlier period, interviewers asked, "During the past 12 months (that is, since (date) a year ago), ABOUT how many days did illness or injury keep [person] in bed all or most of the day? (Include the days in the past 2 weeks.) (Include the days while a patient in a hospital.)" For the later period, interviewers asked, "During the past 12 months, (that is, since 12-month date a year ago), ABOUT how many days did illness or injury keep [person] in bed more than half of the day? (Include days while an overnight patient in a hospital.)" Comparability may be slightly compromised by the specification of "all or most of the day" versus "more than half the day" and "while a patient in a hospital" versus "while an overnight patient in a hospital."

For 1976-1981, follow-up questions encouraged respondents to supply answers falling into broad intervals. Specifically, interviewers asked, "Was it more than 7 days or less than 7 days? Was it more than 30 days or less than 30 days? Was it more than half the year or less than half the year?" The values for BEDAYR2--None, 1-7 days, 8-30 days, 31-180 days, 181-365--mirror the intervals mentioned in these follow-up questions. Beginning in 1982, the follow-up questions were dropped, and respondents had to supply a single number falling between 0 and 365 (for 1982-1996) or 0 and 366 (for 1997 forward). To facilitate comparisons over time, BEDAYR2 recodes the more detailed responses for 1982 forward into the broad intervals that are the only information available in the public use files for 1976-1981. Researchers who are only interested in data from 1982 or later may wish to analyze the non-intervalled data on bed days during the preceding 12 months, included in BEDAYR.

There was a significant change in the variable universe, with all persons included up through 1996 and only sample adults included for 1997 forward. Researchers may wish to restrict the data for BEDAYR2 to persons age 18 and older when comparing results before and after 1997.

Variable Codes and Frequencies

BEDAYR2
Bed disability days in past 12 months, intervalled

For more information about BEDAYR2, see the variable description

An 'X' indicates the category is available for that sample.

Case-Count View

	Code	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	78	77	76	
NIU	0	X	X	X	X	X	X	X	X	X	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	NIU
None	1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	None
1-7 days	2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	1-7 days
8-30 days	3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	8-30 days
31-180 days	4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	31-180 days
181-365 days	5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	181-365 days
Unknown	9	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Unknown

Category Availability View

	Code	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988	1987	1986	1985
NIU	0	67221	63134	61296	62342	67434	68244	66258	66345	67361	·	·	·	·	·	·	·	·	·	·	·	·
None	1	20008	19729	19811	19799	20538	20057	19223	19795	21397	35920	57811	63730	59284	70743	66439	66227	63753	67380	67690	33586	50453
1-7 days	2	8488	8607	7974	8365	9383	9103	8757	9370	11012	20817	33893	39757	38278	43697	40557	40273	39827	40950	40904	20904	30048
8-30 days	3	1652	1655	1617	1501	1838	1872	1649	1882	2153	4084	6774	8285	8335	9447	8870	8983	9221	9574	9923	5372	7782
31-180 days	4	603	637	651	609	678	629	549	722	817	1243	1910	2453	2311	2658	2439	2462	2474	2537	2651	1376	2041
181-365 days	5	162	115	175	164	176	153	119	170	207	316	601	668	579	745	669	648	602	646	637	318	490
Unknown	9	515	583	624	606	713	560	504	501	530	1022	1478	1286	884	1122	1058	1038	1052	1223	1054	496	717

Variable Codes (Composite Coding)

INTEGRATED HEALTH INTERVIEW SERIES

HOME | DATA | DOCUMENTATION | VARIABLES | CONTACT

RACEA
Main Racial Background (Pre-1997 Revised OMB Standards), self-reported or interviewer reported

For more information about RACEA, see the variable description

An 'X' indicates the category is available for that sample.

Case-Count View

	Code	05	04	03	02	01	00	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	78	77	76	75	74	73	72	71	70	69
White		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Black/African-American		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Aleut, Alaskan Native, or American Indian										x	x					x	x	x	x	x	x	x	x	x	x	x	x	x	x	x							
Alaskan Native or American Indian		x	x	x	x	x	x	x	x																			x										
Alaskan Native/Eskimo								x	x	x	x																											
Aleut								x	x	x	x																											
American Indian								x	x	x	x																											
Asian or Pacific Islander														x	x	x	x	x	x	x	x	x	x	x	x	x	x											
Asian																																						
Chinese		x	x	x	x	x	x	x	x	x	x	x	x	x	x																							
Filipino		x	x	x	x	x	x	x	x	x	x	x	x	x	x																							
Korean											x	x	x	x																								
Vietnamese											x	x	x																									
Japanese											x	x	x																									
Asian Indian		x	x	x	x	x	x	x	x	x		x	x	x																								
Pacific Islander																																						
Hawaiian											x	x	x																									
Samoan											x	x	x																									
Guamian											x	x	x																									
Other Asian or Pacific Islander																																						

Data Extract Preferences

Data Extract Variable Selection

General Health Variables (P) [top]

Variable Name	Variable Label	05	00	95	90	85	80	75	70
☐ BDDAY2WK	Bed disability days in past 2 weeks	.	.	X	X	X	X	X	X
☐ BEDAYR	Bed disability days, past 12 months	X	X	X	X	X	.	.	.
☐ BEDAYR2	Bed disability days in past 12 months, intervalled	X	X	X	X	X	X	.	.
☐ BMI	Body mass index	X	X	.	.	.	.	.	.
☐ BMICALC	Body Mass Index, calculated from publicly released height and weight variables	X	X	X	X	X	X	.	.
☐ HEALTH	Health status	X	X	X	X	X	X	X	.
☐ HEIGHT	Height in inches without shoes	X	X	X	X	X	X	.	.
☐ HSTATYR	Health status compared to 1 year ago	X	X	.	.	.	.	.	.
☐ RADAY2WK	Other restricted activity days in past 2 weeks	.	.	X	X	X	X	X	X
☐ RADAY2WKRSID	Other restricted activity days in past 2 weeks - Residual Measure	.	.	X	X	X	.	.	X
☐ RADAY2WKTOT	Total restricted activity days in past 2 weeks	.	.	X	X	X	X	X	.
☐ SLDAY2WK	School loss days in past two weeks	.	.	X	X	X	X	X	X
☐ SLDAYR	School loss days, past 12 months	X	X	.	.	.	.	.	.
☐ WEIGHT	Weight in pounds without clothes or shoes	X	X	X	X	X	X	.	.
☐ WLDAY2WK	Work loss days in past 2 weeks	.	.	X	X	X	X	X	X
☐ WLDAYR	Work loss days, past 12 months	X	X	.	.	.	.	.	.

Medical Care Variables (P) [top]

Variable Name	Variable Label	05	00	95	90	85	80	75	70
☐ CARE10X	Received care 10+ times in past 12 months	X	X	.	.	.	.	.	.
☐ CHECKUP	Had checkup in past 12 months	X	X	.	.	.	.	.	.

Data Extract Summary

Download Data & Documentation

IHIS Data Releases

- First public data release: **November 2006**

- Second public data release: **June 25 2007**
  - ☐ Coincided with 50th Anniversary Celebration of NHIS

- IHIS now includes over 1000 integrated variables from:
  - ☐ Household and Person files for 1969-2006
  - ☐ Sample Adult/Sample Child files for 1997-2006
  - ☐ Access to Care supplements 1993-1996

NHIS-IHIS Linking Keys

- **Unique identifiers for each person and household within each survey year**
  - ☐ A single variable, which is comparable across years
  - ☐ Constructed from a series of data elements in the NHIS
  - ☐ Consistent with NHIS unique identifiers

- **Allows data from the original NHIS public use files to be linked to IHIS data files**
  - ☐ In cases where a variable is not yet in IHIS

- **User note provides guidance on linking**

- **Stata and SAS syntax files provided to facilitate linking process**

Future IHIS Data Releases

- Additional IHIS data will be released to the public once or twice per year

    - Accommodate other units of analysis, such as conditions, doctor visits, hospital episodes

    - Incorporate harmonized variables from other NHIS data files*

    - Include soon-to-be released data from newly created NHIS public-use files from 1963 – 1968

    - On-line questionnaire text linked to the corresponding variable descriptions

    - On-line tabulation capability, variable search engine, and improved user interface available by early 2008

For Additional Information

INTEGRATED HEALTH INTERVIEW SERIES

- http://www.ihis.us
- ihis@pop.umn.edu

SHADAC — STATE HEALTH ACCESS DATA ASSISTANCE CENTER

- http://www.shadac.umn.edu

MPC MINNESOTA POPULATION CENTER

- http://www.pop.umn.edu