

等で貴重なコメントをいただいた。この場をかりて御礼を申し上げたい。
なお、本稿は、平成8年度『所得再分配調査』の調査票を、厚生省より許可を得て使用したものである（平成12年1月7日統発第2号）。
本稿の見解は、筆者個人のものであり、筆者の所属する組織の見解を表したものではない。

参考文献：

- 阿部 彩(2000)「社会保険料の逆進性が世代内所得不平等度にもたらす影響」『季刊社会保障研究』36巻1号、pp.67-80。
小椋正立・角田 保(2000)「世帯データによる社会保険料負担の納付と徴収に関する分析」『経済研究』Vol.51, No.2, pp.97-110。
厚生省(1999)『厚生白書(平成11年版)』。
厚生省大臣官房統計情報部『平成8年国民生活基礎調査』
社会保険庁(1997a)『平成7年公的年金加入状況等調査報告』。
社会保険庁(1997b)『平成8年国民年金被保険者実態調査』。
社会保険庁(2000)『平成10年公的年金加入状況等調査報告』。
鈴木 亘・周燕飛(2000)「国民年金未加入者の経済分析」大阪大学社会経済研究所 DP No.508。
八木匡・橋木俊詔(1996)「等価所得比率の測定と所得分配不平等度の解釈」『季刊社会保障研究』32巻2号、pp.178-189。

付録1 国民年金の保険料設定と免除制度

国民年金保険の保険料は、被保険者の収入いかに係わらず月額 13,300 円(1995 年1月～3月 11,000 円、4月～12月 11,700 円)の定額であるが、低所得者に対しては免除制度が設けられている。免除には2つの種類があり、一つは法定免除といい、障害を支給事由とする年金の受給権者、生活保護受給者、施設入所者が自動的に対象となる(国民年金法第89条)。もう一方は、所得がないとき等一定の事由に該当する場合やその他保険料を納付することが著しく困難であると認められる場合に適用され、被保険者は、市町村による審査の上、免除を受けることができる。これを申請免除という(同法第90条)。申請免除の対象は、所得のない者、被保険者又は世帯の他の世帯員が生活扶助以外の扶助を受ける者、地方税法に定める障害者又は寡婦であり年間の所得が125万円(1995年130万円)以下である者、その他保険料を納付することが著しく困難であると認められるものである。
さらに、厚生労働省は、次のように申請免除の基準を定めている。

国民年金の保険料免除(申請免除)の基準

一般免除基準

- ① 被保険者、その世帯主・配偶者に前年分の所得税額があるときは免除せず、当該年度分の市町村民税が賦課されていないときは免除する。
- ② ①によって決定できない場合には、世帯全員の前年の所得額を基に一定の方法によって算定した指数により免除の可否を決定する。
- ③ ①および②の結果、免除に該当しない場合でも、失業、倒産、その他の理由で申請時の所得状況等が前年度の所得状況と著しく異なる等により、保険料の納付が困難と認められるときには免除できる(以下「一定特例免除」という)。
- ④ ①ないし③により免除に該当する場合であっても、高額(7万円以上)な生命保険等の保険料を払っている場合にはその支払額相当の保険料負担能力があるものとし、また、①及び②により免除に該当する場合であっても、著しく高額(14万円以上)な生命保険料等を支払っている場合には、免除しないことができる。

学生に係わる保険料免除基準

- ①本人に前年分の所得税額がある場合は、免除しない。
- ②本人に前年分の所得税額がない場合は、親元の世帯員全員につき、前年の所得に基づき計算を行い、その結果が次表の金額を下回る場合は免除とする。
- ③②により免除に該当するものでも、前年度に比較して所得が著しく改善した場合は免除しない。
- ④①ないし②により免除に該当しない場合でも、学生本人又は親元世帯が次に該当する場合は生活程度等を考慮して免除とすることができる。ア.前・当年度に災害により著しい損害を受け、保険料の拠出が困難、イ.失業・倒産等で、申請時の所得が前年度と著しく相違し、保険料の拠出が困難。(総務庁行政監察局1998)

1998年度の免除率の内訳は、法定免除4.5%、申請免除15.4%である(社会保険庁)。時系列で見ると、法定免除の割合は、4.5%前後で横ばいなのに対し、申請免除の割合は年々増加している。免除制度の運用は市町村によってばらつきがあり、特に一般特例免除に関しては、現場の個別判断に任されており、免除理由の立証資料の提出も求められていない(総務庁行政監察局 1998)。このため、免除制度の恩恵を受けている被保険者の、実質の負担能力はわからないのが実状である。

付録2 未加入、未納、免除のタイプの判断

『再分配調査』からは、支払った保険料の額はわかるが、その個人が未納なのか、免除を受けたのか、またその場合、どのタイプの免除を受けたのかはわからない。そのため、以下の基準をもって未加入、未納、免除のタイプを判断した。未加入・加入は、調査票にある「公的年金の加入状況」の情報に基づいた。これは調査時点(平成8年6月)での加入状況であり、保険料額は前年のものであるため、保険料と加入状況は必ずしも一致しないが、前年の情報がないため代替とした。

普通免除と判定免除は、それぞれの基準に照らし合わせて判別した。一般の場合は、被保険者、その世帯主・配偶者の所得税・住民税の課税状況によって免除の基準が定められているが、ここでは「世帯」の課税状況で代替とした。指数で判定される場合は、指数の計算に必要な情報が揃わないため、全員免除されたと仮定した(所得税非課税、住民税課税であり、保険料を全く払っていないもの)。学生の場合は、本人の所得税課税状況で、まず判断し、免除却下とならない場合は、親元の世帯所得や税控除額、学生人数をもって判定される。ここでは、親・親族と同居している学生は、同居世帯の所得が195万円(同居の場合の免除最高額)、学生本人が世帯主の場合は、本人所得が235万円(別居の場合の免除最高額)以下のものを免除、それ以外のものを免除却下と判断した。最後に、免除とならず、保険料を1円も払っていないものは、未納か特例免除となるが、特例免除とする情報がないため、全員未納とした。

これらの定義に基づき、『平成8年度所得再分配調査』のデータの未加入率、未納率を計算すると、前述の公的調査より大幅に高い値となる。その理由は、『再分配調査』が低所得者層のサンプルが多いこと、他調査(『実態調査』)は郵送調査であり未納者の回収率が納付者のそれより低いことなどが考えられる。本稿では、このため、制度改正のシミュレーションの際には、現行の制度による推計値と制度改正後の推計値の変化率のみを議論の対象とする。

要旨

本研究の目的は、1999年国民年金改正に盛り込まれた保険料の免除制度の改正が、未加入者、未納者の行動に与える影響と、国民年金保険料の逆進性に与える影響を、厚生省『平成8年所得再分配調査』の個票を使って推測することである。

本稿では、未加入・未納の行動モデルを、加入・未加入の Selection を伴う、納付・未納の Probit モデルを用いた推計を行った。結果として、まず、未加入と未納の決定要因は構造的に違うことが認められた。保険料率は、未加入に対して大きな影響力が認められないが、未納に対しては大きな影響力が認められた。また最低加入期間が25年であることの加入行動への影響は、検証することができなかった。これらの結果を基に、免除制度の改正の影響を推計すると、未加入率は殆ど変化せず、未納率は改正前に比べ10%前後減少すると予測された。また、保険料体系の逆進性は改善する可能性が高いと予測された。

表1 未加入の理由 (『公的年金加入状況等調査』から)

	平7	平10
制度の未周知	46.2%	41.9%
加入したくない	53.8%	58.1%

	平成7 (%)	平成10 (%)
保険料を払うことが経済的に困難	14.7	15.6
年金制度の将来が不安	8.1	15.5
保険料が高いから	3.7	4.5
貯蓄や個人年金の方が得だから	4.9	3.7
支払う保険料に比べ受給額が低い		3.0
老後も働くつもりだから	2.6	2.8
今から加入しても受給権がない	3.0	2.5
学生であり親に迷惑をかけたくない	4.1	1.4
手続きが面倒	0.2	0.9
年金額が不満	1.7	0.7
他に収入のあてがある	1.1	0.6
まだ若いから	2.9	0.5
制度がよくわからない	0.9	0.4
今までの加入期間で受給権を得られる	0.5	0.2
その他・不詳	5.2	5.8

	平成7 (%)	平成10 (%)
加入の届出は必要ないと思った	14.9	15.9
届ける暇がなかった	8.3	8.0
うっかり届出を忘れていた	7.2	3.8
制度のしくみを知らなかった	15.8	14.2

出典：社会保険庁 平成7年、平成10年「公的年金加入状況等調査」

表2 未納の理由 (『平成8年国民年金被保険者実態調査報告』から)

	主要回答 (%)	複数回答 (%)
総数	100.0	154.6
うっかりして、忘れた	3.6	5.1
保険料の支払い方法が面倒	1.3	3.1
後でまとめて払おうと思った	4.0	7.5
保険料が高く、経済的に払うのが困難	55.4	63.8
学生であり親に負担をかけたくない	6.9	10.2
国民年金をあてにしていない	20.8	41.7
制度の将来が不安	13.7	29.0
個人年金に入っているから	2.6	3.9
自分で働く	1.5	3.4
財産がある	0.1	0.2
その他・不詳	3.0	4.9
まだ若いので今から払わなくてもよいと思う	0.8	3.2
これから保険料を払っても加入期間が足りない	3.0	9.9
すでに厚生年金や共済組合の年金の受給権がある	1.3	2.1
すでに国民年金の受給権がある	0.5	1.6
特に理由はない	2.2	6.4

出典：社会保険庁『平成8年国民年金被保険者実態調査』

表3 記述統計量

	想定第一号被保険者 全サンプル (n=2,814)		最多所得者 (n=1668)		その他の世帯員 (n=1,146)	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
保険料率	0.04017	0.03960	0.03907	0.03406	0.04178	0.04647
性別	0.4360	0.4960	0.2734	0.4458	0.6728	0.4694
年齢(歳)	40.9	11.8	43.4	10.8	37.4	12.3
20~24歳ダミー	0.1311	0.3376	0.0689	0.2534	0.2216	0.4155
25~29歳ダミー	0.1116	0.3149	0.0875	0.2827	0.1466	0.3539
30~34歳ダミー	0.0785	0.2690	0.0767	0.2663	0.0812	0.2732
35~39歳ダミー	0.0896	0.2856	0.1007	0.3010	0.0733	0.2607
40~44歳ダミー	0.1251	0.3309	0.1343	0.3411	0.1117	0.3151
45~49歳ダミー	0.1855	0.3888	0.2014	0.4012	0.1623	0.3689
50~54歳ダミー	0.1350	0.3418	0.1813	0.3679	0.0969	0.2959
55~59歳ダミー	0.1436	0.3507	0.1691	0.3749	0.1065	0.3086
一般常雇者・役員・その他ダミー	0.2676	0.4428	0.3183	0.4660	0.1937	0.3954
自営業者ダミー	0.4929	0.5000	0.6174	0.4998	0.4572	0.4984
契約被雇用者ダミー	0.0757	0.2646	0.0659	0.2483	0.0899	0.2861
無職ダミー	0.1336	0.3403	0.0743	0.2624	0.2199	0.4144
個人総所得(万円)	247.96	371.52	369.9	432.0	70.5	119.3
等価世帯所得(万円)	377.82	330.90	370.2	315.6	389.0	351.9
生命保険加入ダミー	0.4613	0.4986	0.5989	0.4903	0.2609	0.4393
個人年金加入ダミー	0.1272	0.3333	0.1469	0.3541	0.0986	0.2983
居住地規模 大都市ダミー	0.1759	0.3808	0.1882	0.3910	0.1579	0.3648
人口15万以上の市ダミー	0.2886	0.4532	0.2890	0.4534	0.2880	0.4530
人口5万以上15万以下ダミー	0.2356	0.4245	0.2338	0.4234	0.2382	0.4262
人口5万以下の市ダミー	0.0675	0.2510	0.0719	0.02585	0.0611	0.2396
郡部	0.2324	0.4224	0.2170	0.4123	0.1937	0.3954

想定第一号被保険者 = 第一号被保険者 + 未加入者

保険料率 = 想定保険料 / 世帯等価所得

性別 = 男性0、女性1

契約被雇用者 = 1年未満の契約または家庭内職者

等価世帯所得 = 世帯所得 / 等価所得比率(世帯人員数)

表4 Probit with Selection Model 推計結果

被説明変数 説明変数	モデル1(全サンプル)		モデル2(最多所得者)		モデル2(それ以外の世帯員)	
	係数	標準誤差	係数	標準誤差	係数	標準誤差
納付(Probit 部分)						
保険料率	-9.2184 ***	0.8819	-7.6495 ***	1.1031	-11.8402 ***	1.6182
性別	-0.2896 ***	0.0539	-0.2406 ***	0.7707	-0.1125	0.0940
年齢	0.0202 ***	0.0027	0.0195 ***	0.0035	0.0164 ***	0.0045
生命保険加入	0.3998 ***	0.0591	0.2591 ***	0.0742	0.5132 ***	0.1076
個人年金加入	-0.0945	0.0806	-0.1121	0.0986	-0.0581	0.1423
居住地規模 大都市*	-0.1567	0.0896	-0.2086	0.1139	-0.1845	0.1556
人口15万以上の市	-0.1956	0.0763	-0.1943	0.1006	-0.2648 *	0.1236
人口5万以上 15 万以下	-0.2357 ***	0.0773	-0.2649 *	0.1034	-0.2573 *	0.1191
人口5万以下の市	-0.0180	0.1146	-0.0906	0.1483	0.0129	0.1833
定数項	-0.0358	0.1278	0.0908	0.1778	-0.0009	0.1911
加入(Selection 部分)						
保険料率	-0.4179	0.7032	-0.6913	1.0567	-0.3222	0.9638
性別	0.1329 *	0.0601	0.1194	0.0875	0.0272	0.0985
年齢層						
20代後半	0.3299 ***	0.0999	0.3410 *	0.1629	0.3268 *	0.1293
30代前半	0.4064 ***	0.1159	0.4774 ***	0.1790	0.3224 *	0.1628
30代後半	0.4734 ***	0.1180	0.3662 *	0.1674	0.6276 ***	0.1884
40代前半	0.4322 ***	0.1105	0.2752	0.1612	0.7285 ***	0.1763
40代後半	0.3105 ***	0.0984	0.2095	0.1458	0.5173 ***	0.1557
50代前半	0.3762 ***	0.1109	0.2997	0.1583	0.5505 ***	0.1890
50代後半	0.5926 ***	0.1147	0.4990 ***	0.1605	0.8269 ***	0.2073
無職	-0.3595 ***	0.0814	-0.4783 ***	0.1271	-0.2769 *	0.1185
自営業者	0.4171 ***	0.0700	0.4508 ***	0.8793	0.3396 ***	0.1224
契約従業者	-0.1664	0.1009	-0.1387	0.1400	-0.2126	0.1509
生命保険加入	0.1508 *	0.0644	0.1202	0.0799	0.3601 ***	0.1275
個人年金加入	0.0489	0.0996	0.1132	0.1192	-0.1137	0.1887
居住地規模 大都市*	-0.5774 ***	0.0952	-0.5708 ***	0.1270	-0.5962 ***	0.1472
人口15万以上の市	-0.4351 ***	0.0879	-0.4720 ***	0.1195	-0.3866 ***	0.1320
人口5万以上 15 万以下	-0.2255	0.0935	-0.2950 *	0.1263	-0.1292	0.1416
人口5万以下の市	-0.1679	0.1383	-0.1578	0.1805	-0.2250	0.2199
定数項	0.6981 ***	0.1105	0.7739 ***	0.1663	0.6732 ***	0.1597
Log Likelihood	-2438.68		-1504.066		-1011.199 -2515.265	
ρ	-0.73511 ***	0.0938	-0.6933 ***	0.1250663	-0.7768 ***	0.1395
サンプル数	2814		1668		1146	

*** 1%有意 * 5%有意

ダミー変数の基準個人 性別=男性、年齢階級=20歳代前半、職業=一般常雇者・役員・その他
居住地=郡部

表5 免除制度改正前と後の対象者数 (人数)

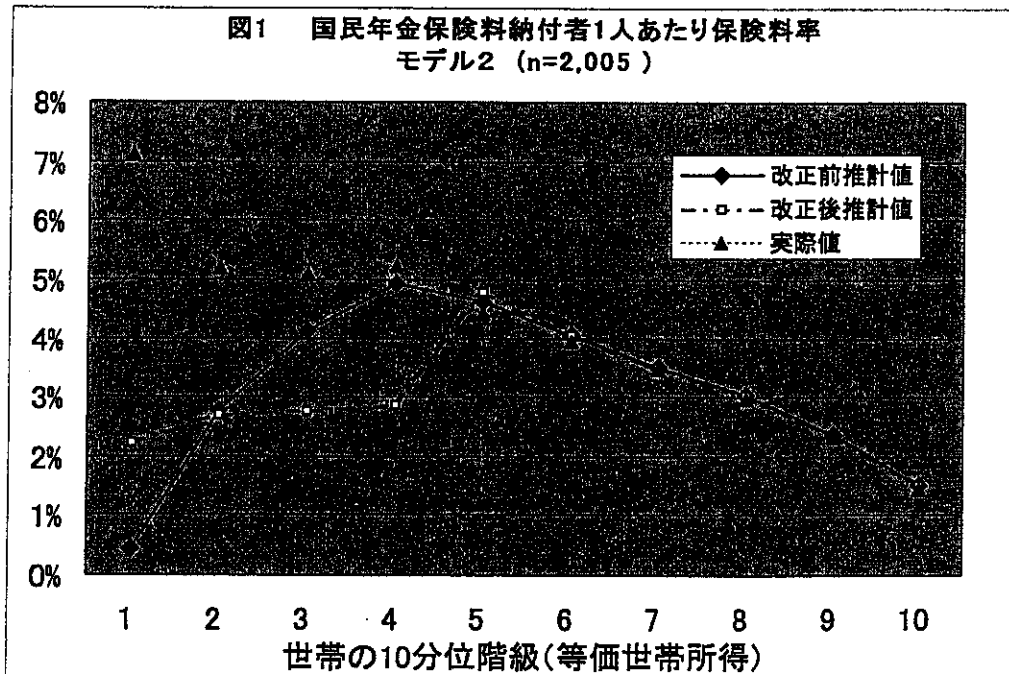
合計 (計 2,814 人)

改正前	改正後			計
	全額免除	半額免除	免除なし	
全額免除	411 (14.6%)	59(2.1%)	10(0.36%)	480
免除なし	0(0%)	680(22.4%)	1704(60.6%)	2334
計	411	689	1714	2814

表6 免除制度改正が与える影響：変化率の推計

	モデル1	モデル2		
		最多所得者	それ以外の世帯員	合計
未加入率変化率	-0.04%	0.00%	0.00%	0.00%
未納率変化率	-11.63%	-7.77%	-12.60%	-9.92%
保険料収入	0.98%	-6.34%	9.01%	-2.58%

図1



The LIS Project: Overview and Recent Developments

Timothy M. Smeeding
Overall Project Director
Luxembourg Income Study
and
Director
Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, NY 13244-1020
(315) 443-9042 | fax (315) 443-1081
tmsmeed@maxwell.syr.edu

Prepared for the
"Japanese Journal of Social Security Research Abroad"

March 2001

Key Words: Luxembourg Income Study (LIS); harmonization, comparability, remote access, infrastructure, household income data, inequality, poverty

*The author wishes to thank the LIS country sponsors for their support in producing this paper. Esther Gray, Mary Santy, and Kati Foley provided excellent assistance in preparing the manuscript. The paper has also benefited from the comments of David Jesuit and Paul Alkemade. However, the author assumes responsibility for any and all remaining errors of omission and commission.

Abstract

The Luxembourg Income Study (LIS) project is one of the oldest and best known examples of cross-national social science infrastructure. Some 25 nations and 20 financial sponsors team together to provide internet accessible, privacy-protected, household income microdata to over 400 users in 30 nations. The project is financed by annual contributions by National Science Foundations and/or National Statistical Offices. One of the most crucial pieces of the LIS project is the source and type of data that it offers to its users. This paper describes these data, how they are obtained, harmonized, and made available. It also presents a critical discussion of recent findings, future directions, and where and how international data collection efforts can improve upon both the quality of income data and its dissemination to qualified researchers.

I. Introduction

The Luxembourg Income Study (LIS) research and databank project has provided harmonized cross-national household income microdata for social science research for over 15 years. These data provide the basis for cross-national comparative research projects by providing access to household income microdata for all research users who are connected to the internet, who promise to respect the privacy of survey respondents, and who promise to make use of the LIS microdata for research purposes only.

The purpose of this paper is to describe the types of data used by LIS and the issues involved with obtaining, harmonizing, and making the data available to users. I also describe recent findings and future research directions. We begin with a description of LIS and the types of data it employs. We then turn to a more in-depth discussion of data type, data quality, and cases of “typical” data from a subset of nations. We discuss additional cases in which microdata have not yet been obtained, and dilemmas regarding privacy protection for data that have been made available to LIS. Next we add recent research findings and then close with a brief view of future LIS plans. The objective here is to give the nonuser a brief overview of the data sources used by LIS and the way that they are harmonized, deployed, and accessed in a time-tested privacy-protected manner by over 400 users in 30 nations, 24 hours a day, 365 days a year. It is also to provide an overview of the research findings of LIS and its future research directions.

II. The Luxemburg Income Study: A Brief Overview

The Luxembourg Income Study (LIS) project began in 1983 under the joint sponsorship of the government of Luxembourg and the Center for Population, Poverty, and Policy Studies (CEPS) in Luxembourg. The LIS project has five goals:

- to *harmonize* cross-national data (through relieving researchers of this task) and by building an expert staff to accomplish this task and to handle user questions and user services;
- to test the feasibility of *creating* a database consisting of social and economic household survey microdata from different countries:
- to provide a method of allowing researchers to *access* these data under various privacy restrictions required by the countries providing the data;
- to create a system that will allow research requests to be *quickly processed* and the responses returned to users at remote locations; and
- to *promote comparative research* on the economic and social status of populations in different countries, through training and networking activities.

CEPS, which later became CEPS/INSTEAD (International Networks for Studies in Technology, Environment, Alternatives, Development, or INSTEAD) was important because its leader, Gaston Schaber, understood the importance of cross-national data and cross-national comparative research studies for the tiny country of Luxembourg, which otherwise would attract very little international scholarly attention. He saw the production of harmonized (comparative) cross-national data as a key step forward in the social sciences. Thus, he funded the beginning of the LIS project, with the author as project director, with Lee Rainwater at Harvard University as research director, and with John Coder of the U.S. Census Bureau as technical director. Also important to LIS was the Luxembourgish government, whose data protection laws are very strict, thus allowing guaranteed privacy protection for household microdata to national statistical offices who had no experience with making data available outside of their national boundaries.

From the beginning, the LIS project was supported by a key group of academics and social statisticians who were valuable because of their intellectual capital and their ability to make datasets and technical expertise available to LIS.¹ LIS produced its first research outputs in 1985, and became fully available to researchers in 1987. In terms of general access, therefore,

it is about 14 to 15 years old. It stands as one of the few truly cross-national and comparable data infrastructures extant (OECD 2000), and thus provides a useful model for other similar projects.

Financing a Public Good

The project is now funded on a continuing basis by the national science foundations and social science research foundations of its member countries, by the ingenuity of its directors, and by the Luxembourg government through CEPS/INSTEAD, a government agency within which LIS is located. The original grant from the government of Luxembourg in 1983 was followed by a "bridge grant" from the Ford Foundation, allowing LIS the time to find sponsoring partners in most of its member nations. By 2000, 16 nations help support the LIS core via 24 different funders, each of whom pay between \$10,000 and \$30,000 per year (see Table 1 for a list of these sponsors). The list is a broad one: central statistical offices, national science foundations, social science foundations, and national research institutes. These nations cover, in an average year, over 90 percent of the core costs of the LIS database. Various foundations in the United States (e.g., Ford Foundation, National Institutes of Health, National Science Foundation, Russell Sage Foundation, MacArthur Foundation) and foreign sponsors (European Community) have helped us fill in the funds to meet project demands that exceeded national contributions. For instance, the Ford Foundation covers the cost of maintaining the United States office of LIS at Syracuse University. LIS pays its own expenses except for office rent, which is covered by CEPS/INSTEAD. Because of some previous shortfalls in national contributions, the directors of LIS spend about one-third of their time on the LIS project working on research projects which will ultimately fund the LIS core budget, and which will add new dimensions to the database.

The goals of LIS have in effect made it a "public good." Once the data are harmonized, they are made available to qualified academic users at zero marginal monetary cost. Moreover,

LIS holds summer workshops and other seminars aimed explicitly at increasing the base of users, especially among junior scholars. Over 500 scholars have attended these workshops, many of whom have become long term LIS users. All of these students have been treated to an intellectually rich and academically strenuous opportunity to learn to use the data.

The problem with public goods, however, is that they can be used without helping covering their fixed costs. Within nations, public goods are funded by national bodies and national research funds. But across nations there are few, if any, organizations with the scope or interest to fund a microdata infrastructure (OECD 2000).

As with all real public goods, LIS is therefore, almost continually underfunded. Major international research organizations (e.g., Office of Economic Cooperation and Development (OECD), International Labor Office (ILO), United Nations (UN), World Bank, United Nations Children's Fund (UNICEF), and dozens of others) have followed economic logic and have basically become free riders, making use of LIS data and user services at zero marginal cost. Occasionally, these organizations will hire LIS to provide custom-made analyses which provide some support for LIS. More often, they will hire one or more of the 400 currently active LIS users at marginal cost to carry out these analyses. The result is that LIS has continually been faced with "projectizing" the core...adding "data costs" to research projects so that the core could be maintained, updated, and improved.

Yet, country contributions have been amazingly steady. In one country (France), the sponsoring organizations which send checks to LIS were disbanded and thus we have had to find new support. In some nations there are lapses in sponsorship (e.g., Spain, Austria). In all other nations, support has not only continued but has been increased when asked. Most of the sponsors listed in Table 1 have been LIS contributors for over ten years. However, raising the core country membership price will, at some point, reduce membership. Thus, we have not until

very recently been able to totally fund LIS from the core alone and future core funding may be an issue.

Should country contributions and core staff costs continue at current levels, we will for the first time be able to fund the core from country revenues alone. In this situation, we can turn our attention directly to improving the core: data quality, user access and user services, substituting these activities for the time intensive work of projecting the core.

Data Harmonization

The most important goal for LIS, we believe, is data harmonization. Data availability is an important issue that is slowly being overcome (see sections III and IV below). But the access and availability of 3, 4, or more national income surveys with no idea of how sources or definitions of “income” are arrived at does not permit comparability. Harmonization of data—reshaping and reclassifying components of income or definitions of household structure into comparable categories—is the real value of LIS. It allows the researcher to address important social issues without having to invest countless hours getting every variable that will be analyzed into a comparable format.

Because of data restrictions and privacy concerns of many governments, LIS must keep the data in one location where it can be accessible yet “protected” against misuse. The LIS micro datasets are therefore accessed globally at zero direct cost to their user using electronic mail. More general release of LIS data to national archives is difficult due to differential national interests in data protection for clients and governments (e.g., Japan, Sweden, Finland, others); sale of national data to recover costs (e.g., Canada, Australia, the United Kingdom, others), and other complicated political prerogatives (e.g., the European Community Household Panel Dataset (ECHP)), all of which are described more fully below in sections III and IV. Despite these issues, national or international statistical bodies which would like to make data

available but also protect privacy and confidentiality ought to consider LIS or similar organizations as a method of providing access to their data at reasonable cost and with no risk of violating the confidentiality and privacy of survey respondents.

Countries Covered and Access

Since its beginning, the LIS experiment has grown into a cooperative research project with a membership that includes countries in Europe, North America, the Far East, and Australia. Our countries are largely covered by the OECD, G-7, and in the European Community broadly defined. The database now contains information for almost 30 countries for one or more years of data. Negotiations are underway to add data from New Zealand, Korea, Japan, South Africa, and other countries. The LIS data bank includes more than 100 datasets covering the period of 1968 to 1997. During 2001 additional surveys are being added to more fully represent the period of the middle 1990s for most of the nations, and in 2002 we will begin a new "millennium" round of datasets for 2000. A list of countries and years for which data are available is attached (Table 2).

A new operating system for our remote access network was implemented in 1998, and already improved again in 2000. It is much more flexible and faster than the old system, and it provides additional data access options including SAS, SPSS, and STATA software packages. Extensive documentation concerning technical aspects of the survey data and the social institutions of income provision in member countries is also available to users via electronic mail. In 1999 we began to provide direct web-access to "mesodata" and "metadata" in the form of comparable output on income distribution, poverty, and related issues. Inexperienced users can now download these tabular data in comparable form. In future years LIS will add a new "web tabulator" system that allows inexperienced users the ability to obtain summary data by only entering a few key words into a worldwide web-based system which will generate these tabulations directly.

II. Data Details and Case Studies

As seen in Table 3, there are numerous types of data to which LIS has access. Each nation's data is almost a story in and of itself. The various nations follow very different policies with respect to data access, data quality, and data availability. Types of survey data available are listed in Table 3, where we present data by type (3A) and by one measure of dataset quality (3B).

Survey Types and Data Quality

Perhaps the most important issue of comparability lies with the relative quality and consistency of LIS datasets themselves. The types of survey data used by the LIS are not uniform in nature, purpose, or objective. The lowest common denominator the LIS requires is the existence of a substantial level of detail concerning income sources and income totals. The surveys themselves are quite diverse, as illustrated in Table 3A. Some surveys are designed first and foremost to collect income data; others are derived from income tax records; and still others come from special supplements to labor force surveys. Some LIS datasets are based on income questions taken from expenditure surveys (as in the case of the United Kingdom); others are separate waves of longitudinal household panel data from a scientific university or research center based data collection (e.g., Germany, Russia); and still others are taken, at least in part, directly from government administrative data. In many nations, several different types of data are available, allowing LIS to choose the "best" survey for comparability reasons (see Atkinson, Rainwater, and Smeeding 1995).

The second part of Table 3B presents a reasonable way to envision how these differences are likely to affect the quality of income data. Five conceptual levels of income reporting are suggested and the approximate level at which each LIS country dataset lies. Income reporting in the upper rows is considered more complete than in lower rows. This is not intended to be more than indicative. There are studies of income distribution which lie between rows 1 and 2, such as

those combining administrative data with survey information, as in the case of Sweden and Finland (see below) or the Blue Book series in the United Kingdom, which combines survey and administrative data.

Up the rows from bottom to top, Table 3B begins with the amount of income actually reported by the population, excluding entire non-interviews but leaving partial or "item" non-response intact (row 5), as in the case in the Dutch, German, and Swiss surveys. The Dutch and Swiss make extensive imputations for some types of income (e.g., social security, child allowances). The German Socioeconomic Panel Data, however, leaves item non-response as missing values, allowing the user to make further imputations for non-reporting of income items. This income data is perhaps more congruous with the next level, which is edited income (row 4) whereby all item non-responses are corrected. These adjustments may take many forms, including "hot-deck" imputation (e.g., the United States Census Bureau technique), where there is imputation of a value taken from the most recent (on the tape) respondent with the same characteristics as the non-respondent, or "cold deck" imputation, whereby the imputation is taken from a matrix which computes the average value of all respondents and assigns the average value to the non-respondent. (See Atkinson, Rainwater, and Smeeding 1995, Appendix 4 for more detail on this topic.)

Row 3 refers to the amount of income recorded in data taken from tax records. Norwegian and French data are at this level. Table 3B suggests that incomes for tax purposes are more reliably reported than survey incomes, which may be true for some but not all countries. Tax-based surveys may also suffer from omissions of certain types of non-taxable income or non-taxpayers, in addition to tax evasion and tax avoidance. Row 2 raises gross incomes to the total amount recorded by some administrative intermediary, based on totals drawn from national income accounts or administrative records of government agencies. Swedish data, for example,

are mainly drawn from such records. Differences between the top row, "true income," and the administrative amounts usually arise from amounts of income which in principle are recorded in the national accounts, but are not readily allocated to individual households. This largely includes the underground, informal, or "shadow" economy as well as fiduciary accounts such as pension funds. These differences in data quality can manifest themselves as differences in the amount and type of income data collected, an issue on which we can briefly comment.

Similarities and differences in the quality of reported income amounts are important in survey measurement. What can be learned about the overall quality of income data from comparisons with national accounts and other external sources is an important question for the LIS; but one for which there is no firm answer. Three points should be made before comparing reported income amounts from surveys and administrative sources. First, national income accounts or administrative data may not always be superior to survey data in some countries. National accounts aggregates are themselves estimates whose reliability is the subject of much literature. Self-employment income, for example, is poorly reported and differs according to the accounting convention employed by the data tabulator. In the case of property income, which is derived as a residual in National Accounts, estimates may be very suspect.

Second, administrative data need adjusting to produce estimates for comparable income concepts and populations before comparing it to survey data (or tax data). For example, national accounts may include households together with non-profit organizations. It may be necessary to subtract the interest income received by charities, or income received by households not in the survey population (e.g., non-residents, the deceased, and the institutionalized), or payments to institutions.

As put by the U.S. Bureau of the Census,

Deriving independent estimates of aggregate income for purposes of evaluating the survey data is difficult. The survey and administrative sources use different definitions, cover different universes, and are based on concepts that are not exactly the same. Therefore, adjustments to the administrative sources must be made to help correct for these inconsistencies and arrive at a valid independent estimate that can be used to make fair and accurate estimates of the quality of the survey estimates (...). [In the United States, these adjustments] attempt to 1) remove income that is received by the institutional population, the deceased, and persons not residing in the United States at the time of interview, 2) remove any components of income that are received as "in-kind" payments or benefits, and 3) remove any lump-sum or one-time payments, withdrawals, etc. (U.S. Department of Commerce 1991, p. 215).

Third, it is important when comparing income amounts to bear in mind that differences between income aggregates may arise from different sources: varying non-response to the survey (for example, a low response rate from high income groups may cause understated investment income); item non-response by households taking part; or inaccurate reporting by respondents. If reported wages and salaries are, say 95 percent, of the comparable aggregate, this does not mean that all individuals reported 95 percent of their true wages and salaries. This is an average based on some individuals who have over-reported or under-reported their incomes. Multiplying reported amounts by the reciprocal of the percentage reported is not the appropriate way to make an adjustment for under-reporting. A direct record-for-record comparison is needed for further information here. Under-recording may appear as failure to report in income source, but it may be indistinguishable from genuine zero entries, creating another type of dilemma. Overall ratings of data quality do not therefore provide all of the ingredients necessary to adjust microdata for reporting errors. Simple "grossing up" will therefore not improve the accuracy of income reporting, even if it produces a higher (but not a better) reported income amount.

Most of the datasets in LIS conform to a reported amount that is overall 85 to 90 percent of the comparable aggregate among the dozen nations who have made these calculations (e.g.,

see Atkinson, Rainwater, and Smeeding 1995, Table 3.7). Wage and salary income tends to be reported with 95 percent or above accuracy. Self-employment income and income from property (interest rents and dividends) are less well-reported. Income transfers fall somewhere in between. However, until we are able to “exactly” match reported incomes with administrative records for the same persons and units (e.g., Radner 1983) we are unable to thoroughly assess data quality.

The bottom line is that all survey income have some error. The degree of error which is tolerable depends on the purpose to which the data will be used. As reported in Gottschalk and Smeeding (2000), the importance of data quality depends on the ratio of the signal (accurate data) to noise (or spurious data). LIS can improve the ratio of signal to noise by making data more comparable; it cannot improve the quality of the data themselves. Others, e.g., the Canberra Group (see below), can improve data quality directly and are therefore of great interest to LIS.

LIS Criteria for Data Selection

Several considerations go into deciding which survey is “best” for LIS purposes:

- **data quality.** The overriding criteria for inclusion in LIS is that this is the highest quality and most consistent and reliable national dataset for measuring annual household income and its components.
- **income detail.** The more detail on an income survey, the better the estimate of income. In particular, surveys explicitly designed to measure “income” do a better job.
- **national staff support.** Every LIS dataset has one or more national country coordinators, who help with technical documentation, harmonization of data, and with user support that goes beyond the knowledge of the LIS team.
- **periodicity.** In general we now try to have data for most nations on a four to five year period rotating basis. We cannot include every year’s data for every nation due to cost. On the other hand if a nation has only one or two years of “good” data, we will include these years even if they do not closely match to other nations. In general, LIS seeks to “space” datasets first, and second, to find a “given” year, e.g., 1995 or