

## Diabetes Information Technology & WebWatch

# Usage of the [www.2aida.org](http://www.2aida.org) AIDA Diabetes Software Website: A Pilot Study

ELDON D. LEHMANN, M.B. B.S., B.Sc.<sup>1,2</sup>

### ABSTRACT

AIDA is a diabetes-computing program freely available from [www.2aida.org](http://www.2aida.org) on the Web. The software is intended to serve as an educational support tool, and can be used by anyone who has an interest in diabetes, whether they be patients, relatives, health-care professionals, or students. In previous "Diabetes Information Technology & WebWatch" columns various indicators of usage of the AIDA program have been reviewed, and various comments from users of the software have been documented. One aspect of AIDA, though, that has been of considerable interest has been to investigate its Web-based distribution as a wider paradigm for more general medically related usage of the Internet. In this respect we have been keen to understand in general terms: (1) why people are turning to the Web for health-care/diabetes information; (2) more specifically, what sort of people are making use of the AIDA software; and (3) what benefits they feel might accrue from using the program. To answer these types of questions we have been conducting a series of audits/surveys via the AIDA Website, and via the software program itself, to learn as much as possible about who the AIDA end users really are. The rationale for this work is that, in this way, it should be possible to improve the program as well as tailor future versions of the software to the interests and needs of its users. However, a recurring observation is that data collection is easiest if it is as unobtrusive and innocuous as possible. One aspect of learning as much as possible about diabetes Website visitors and users may be to apply techniques that do not necessitate any visitor or user interaction. There are various programs that can monitor what pages visitors are viewing at a site. As these programs do not require visitors to do anything special, over time some interesting insights into Website usage may be obtained. For the current study we have reviewed anonymous logstats data, which are automatically collected at many Websites, to try and establish a baseline level of usage for the AIDA site. For the initial pilot study the analysis was performed from October 1, 2000 to November 1, 2001. The study has yielded an interesting insight into how the AIDA Website is being used. The results also confirm those of

---

<sup>1</sup>Academic Department of Radiology, Barts and The London NHS Trust, St. Bartholomew's Hospital; and Department of Imaging (MRU), NHLI (Imperial College of Science, Technology and Medicine), Royal Brompton Hospital, London, United Kingdom.

<sup>2</sup>Dr. Lehmann is a co-developer of the AIDA diabetes simulator.

AIDA is an independent, noncommercial development which is being made available free-of-charge via the Internet—at a dot org (.org) not-for-profit Website—as a noncommercial contribution to continuing diabetes education.

previous audits based on different self-reported methodologies, confirming, amongst other things, what countries people are visiting from and what operating systems/computers they are using. These analyses have been informative and useful. Given this, it is proposed to repeat the current pilot survey approach on a routine basis, in the future, as a way of monitoring on-going usage of the AIDA Website.

## INTRODUCTION

**A**IDA IS A FREWARE COMPUTER PROGRAM that permits the interactive simulation of plasma insulin and blood glucose profiles for demonstration, teaching, and self-learning purposes. It has been made freely available, without charge, on the World Wide Web as a non-commercial contribution to continuing diabetes education. In the 6+ years since its original Internet launch well over 200,000 visits have been logged to the AIDA Web pages at [www.2aida.org](http://www.2aida.org), and over 40,000 copies of the program have been downloaded, *gratis*. Further copies have been made available, in the past, on diskette by the system developers and from the British Diabetic Association, London, U.K.<sup>1</sup> The AIDA software and its usage, has been described previously in this journal,<sup>2-10</sup> as well as elsewhere in the literature.<sup>11,12</sup>

In previous articles, and "Diabetes Information Technology & WebWatch" columns, various indicators of usage of AIDA have been reviewed,<sup>13,14</sup> and various comments from users of the software have been documented.<sup>15-18</sup>

One aspect of AIDA, though, that has been of considerable interest has been to investigate the Web-based distribution of the program as a wider paradigm for more general medically related usage of the Internet. In this respect it seems important to understand in general terms: (1) why people are turning to the Web for health-care-related/diabetes information; (2) more specifically what type of people are making use of the AIDA software; and (3) what benefits they feel might accrue from using the program. To answer these sorts of questions we have been conducting a series of audits and surveys via the AIDA Website, as well as via the software itself, to learn as much as possible about who are AIDA end users.<sup>13-15</sup> The rationale for this work has been that in this way it should be possible to im-

prove the program as well as tailor future versions of the software to the interests and needs of its users. However, a recurring observation has been that the data collection is easiest if it is as unobtrusive and innocuous as possible. Intuitively it would seem self-evident that response rates would be higher for surveys that are easy to complete. Therefore, for instance, we have had considerable success with a simple survey conducted during the less than 5 min while AIDA Website visitors download the freeware software—with over 7,500 audit responses received to date (3,797 of these so far analysed<sup>13,14</sup>).

We have also had interesting responses to a pilot survey specifically asking downloaders "How do you see yourself maybe making use of the AIDA software?"<sup>15</sup> However, in all these audits<sup>13-15</sup> we have been reliant on downloaders' self-reported answers. People who choose to reply to these questions may represent a self-selected group of AIDA users. In this respect it would be useful to establish how representative their responses are, compared with all AIDA downloaders/users and Website visitors.

One way of confirming what we have learnt to date from our existing surveys<sup>13-15</sup> may be to apply techniques that do not require any visitor or user interaction. There are various monitoring programs that can try and keep track of what visitors are doing at a particular site and log which Web pages they are viewing. As these programs do not require visitors to do anything out of the ordinary, over time some interesting insights into Website usage may be obtained. Since for the most part these programs deal with anonymous data, there is no way of identifying a particular individual user. As a result any issues of patient confidentiality can be avoided, as the methodology is purely dealing with aggregated (epidemiological-like) data.

## METHODOLOGY

For the current study we have reviewed anonymous logstats data, which are automatically collected at many Websites, to try and establish a baseline level of usage at the AIDA sites. For the initial pilot study the analysis was performed on data from October 1, 2000 to November 1, 2001.

A supplementary purpose of this study was to try and validate the data collection recorded in previous AIDA surveys.<sup>13,14</sup> In this respect, as shown in Table 1, we have seen in two consecutive audits that over half of AIDA downloaders report being people with diabetes, with 12–13% being relatives of patients, and another 12–13% being doctors. It is of considerable interest to know what sort of end users are downloading the AIDA program, because only in this way can we tailor future versions of the software to the particular needs of AIDA's users. Given this, it becomes of some importance to establish the accuracy and reliability of the percentages shown in Table 1. It is clearly not possible to directly validate who is using the AIDA software from automatically extracted data from a logstats-based Website usage survey. Nevertheless, as shown in Figure 1 there is considerable overlap between the type of data collected via the two completely separate methods, which does allow comparisons to be made between study methodologies. In this way it is our working hypothesis that we should be able to indirectly validate some of the self-reported data from previous audits and surveys<sup>13,14</sup> using automatically collected data stored in our Website logstats files.

AIDA has been available free-of-charge on the Web since March 1996. However, AIDA only moved to its own dedicated Website facility and had access to its own logstats (Website statistics) at [www.2aida.org](http://www.2aida.org) as from the end of September 2000. Therefore this initial pilot study to establish the feasibility of making use of these type of data for interpretation was run for 1 year from October 1, 2000 to November 1, 2001. Because of a technical problem most of the raw data for the month of December 2000 were lost, and only summary statistics are available. As a result to give raw data from 365 days for actual analysis the chronological period was extended to 13 months for this initial pilot study.

Raw logstats files list all activity [Web pages and graphics requested ("hits")] from a Website server. Figure 2 shows an example of the type of information recorded in these files. Because of all the information that is stored, the raw files are very large. In total the raw logstats files for this pilot study, excluding the month of December 2000, occupied over 290 MB of disk storage space.

Two methods of analysis were trialled for this pilot study. Raw logstats files at most Websites are often analysed on a daily basis to provide an online summary or overview of site usage. For this study these online analyses, summarised on a monthly basis using the free Webalizer Website software (available from [www.mrunix.net/webalizer](http://www.mrunix.net/webalizer)), were reviewed. In addition the raw logstats files were downloaded from the AIDA servers on an approximately monthly basis and analysed all together offline on a personal computer (PC). The offline

TABLE 1. RESPONSE TO QUESTION "WHICH OF THE FOLLOWING CATEGORIES BEST DESCRIBES YOU?" IN TWO PREVIOUS AIDA DOWNLOAD SURVEYS/AUDITS<sup>13,14</sup>

	<i>AIDA v4.0 download survey<sup>13</sup> November 1999–July 2000 (%)</i>	<i>AIDA v4.3 download survey<sup>14</sup> July 2000–March 2001 (%)</i>
Patients with diabetes	56.0	55.8
Relatives of patient	13.5	12.4
Doctors	13.0	12.1
Students	6.5	5.3
Diabetes educators	4.2	6.4
Nurses	1.6	2.7
Pharmacists	1.1	2.5
"Others"/none of the above	4.0	2.8

Downloader self-reported data.

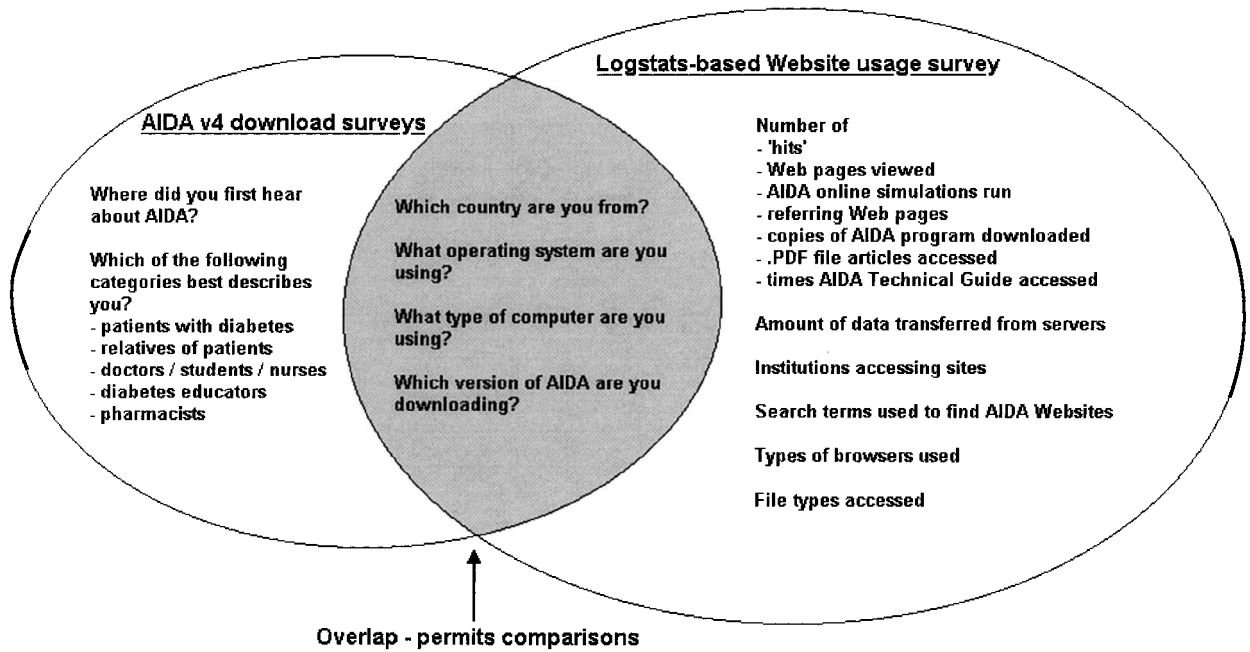


FIG. 1. Schematic diagram overviewing the rationale for the current study, demonstrating the overlap between self-reported data in previous AIDA v4 download surveys<sup>13,14</sup> (left) and automatically collected data in current pilot logstats-based Website usage survey (right). The overlap permits some confirmation of the previously self-reported responses received from AIDA Website visitors. (What type of computer is being used can be inferred, for logstats data, from the operating system that is reported. Therefore, for this study, an Apple Macintosh operating system reported in the logstats file implies an Apple Mac computer is being used.)

analysis was performed using the Analog v4.13 software (University of Cambridge, Statistical Laboratory, Cambridge, UK). Analog is a log-file analysis program that is freely available on the Web (at [www.analog.cx](http://www.analog.cx)) for a wide range of different computer platforms for the analysis of Website logfiles. It is particularly well suited to handling large datafiles. A 266-MHz Pentium II PC was used for the analyses in this study, and this took just 1 min 36 s to summarise and analyse the complete 290+ MB dataset.

An AIDA U.S. mirror site was launched in Baltimore, MD at <http://us.2aida.org> in July 2001. Depending on site traffic, and server status, visitors to the main <http://www.2aida.org> site may be re-routed to the U.S. mirror site (or *vice versa*). In addition, for the most part, downloads of the AIDA freeware software are preferentially handled via the U.S. mirror site, as this has greater bandwidth permitting more downloads to be serviced each month.

For the current pilot study raw logstats data files from the <http://us.2aida.org> and [\[www.2aida.org\]\(http://www.2aida.org\) servers have been analysed together. \(Please note that since this pilot study was completed the AIDA U.S. mirror site has moved from <http://us.2aida.org> to a new separate facility at <http://www.2aida.net> based in California.\)](http://</a></p>
</div>
<div data-bbox=)

## RESULTS

During the 1-year period of this pilot study over 1,522,884 requests for information (or "hits") were logged by the AIDA Website servers. This corresponded with 230,326 actual requests for Web pages. During this time there were on average 4,172 "hits" and 631 Web pages viewed each day. Over the entire 365-day period 40,976 distinct hosts were served, and 10.5 GB of data were transferred, giving an average data transfer rate of 28.7 MB/day.

As is to be expected following the launch of a new Website domain ([www.2aida.org](http://www.2aida.org)), earlier during the survey period (e.g., during October 2000) usage of the Website was less than

```

rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:00 +0000] "GET
/aida/toc.shtml HTTP/1.1" 200 7679 "http://www.2aida.org/" "Mozilla/4.0 (compatible;
MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:01 +0000] "GET
/aida/2intro.shtml HTTP/1.1" 200 3024 "http://www.2aida.org/" "Mozilla/4.0
(compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:01 +0000] "GET
/aida/graphics/new.gif HTTP/1.1" 304 - "http://www.2aida.org/aida/toc.shtml"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:01 +0000] "GET
/aida/intro.htm HTTP/1.1" 200 13924 "http://www.2aida.org/aida/2intro.shtml"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:01 +0000] "GET
/aida/toc2.shtml HTTP/1.1" 200 3974 "http://www.2aida.org/aida/2intro.shtml"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:01 +0000] "GET
/aida/graphics/aida-left.gif HTTP/1.1" 304 - "http://www.2aida.org/aida/intro.htm"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:01 +0000] "GET
/aida/graphics/aida.gif HTTP/1.1" 304 - "http://www.2aida.org/aida/intro.htm"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:01 +0000] "GET
/aida/graphics/aida-right.gif HTTP/1.1" 304 - "http://www.2aida.org/aida/intro.htm"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:01 +0000] "GET
/aida/graphics/bullet.gif HTTP/1.1" 304 - "http://www.2aida.org/aida/intro.htm"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:26 +0000] "GET
/aida/download.htm HTTP/1.1" 200 9679 "http://www.2aida.org/aida/intro.htm"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:27 +0000] "GET
/aida/graphics/home.gif HTTP/1.1" 304 - "http://www.2aida.org/aida/download.htm"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:31 +0000] "GET
/aida/aidadown.htm HTTP/1.1" 200 22619 "http://www.2aida.org/aida/download.htm"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:45:32 +0000] "GET
/aida/graphics/lcompa.gif HTTP/1.1" 200 2694
"http://www.2aida.org/aida/aidadown.htm" "Mozilla/4.0 (compatible; MSIE 5.0; Windows
98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:46:02 +0000] "POST /cgi-
bin/formmail.cgi HTTP/1.1" 302 233 "http://www.2aida.org/aida/aidadown.htm"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:46:05 +0000] "GET
/aida/aidadown2.shtml HTTP/1.1" 200 8929 "http://www.2aida.org/aida/aidadown.htm"
"Mozilla/4.0 (compatible; MSIE 5.0; Windows 98; DigExt)"
rbp2greenfield.cpmc.columbia.edu - - [09/Nov/2000:20:46:58 +0000] "GET
/execs/aidainst.exe HTTP/1.1" 200 1021042
"http://www.2aida.org/aida/aidadown2.shtml" "Mozilla/4.0 (compatible; MSIE 5.0;
Windows 98; DigExt)"

```

FIG. 2. Extract from raw logstats files showing example of data recorded for a visitor from Columbia University. Top entries (in **bold**) refer to loading of AIDA home page. toc.shtml is the left "table of contents" frame bar (menu). 2intro.shtml is a subpage that loads the bottom frame bar (menu)—called toc2.shtml. intro.htm is the actual HTML main AIDA home page. The second block (in *italics*) refers to a visit to the AIDA download page (download.htm). The graphics file lcompa.gif is an image of a computer that appears next to the AIDA software download links. Third block (in **bold**) refers to completion of the AIDA survey/audit on the download page, using an HTML form and the formmail.cgi script (program). The fourth block (in normal typeface) at the bottom of the illustration refers to the downloading of a copy of the AIDA v4.3 executable installation file (aidainst.exe) stored in the [www.2aida.org/execs](http://www.2aida.org/execs) directory. This visitor was using Microsoft Internet Explorer (MSIE) v5.0 with a Windows 98 operating system. The illustration shows the activity at the server on November 9, 2000 between 20:45:00 h and 20:46:58 h GMT.

later (e.g., during October 2001). Figure 3 summarises the number of requests for information (“hits”) during this time. Figure 4 demonstrates how usage of the Website on a day of the week basis shows lower usage over the weekend period (Saturday and Sunday) compared with during the week (Monday–Friday).

Peak usage of the AIDA Website during this 1-year period was recorded at around 22:00 h [Greenwich Mean Time (GMT)], which corresponds with around 6 p.m. Eastern Standard Time and 10 a.m. Pacific Standard Time. The nadir of usage was seen at 06:00 h GMT (Fig. 5).

The Analog software could not resolve domains responsible for 41.3% of traffic (unresolved numerical addresses). However, 17.9% of traffic came from commercial (.com) domains, with 14.3% of traffic coming from network (.net) domains, 8.3% of traffic from .uk (United Kingdom) domains, 2.2% from .edu (U.S. educational) establishments, 0.5% of traffic from .org non-profit-making organisations, and 0.22% from .gov U.S. government establishments, with 0.16% from .mil U.S. military establishments.

In terms of countries accessing the AIDA Websites, visits were logged from 117 countries

(in alphabetical order): Andorra, Argentina, Armenia, Aruba, Australia, Austria, Belarus, Belgium, Bermuda, Bhutan, Bolivia, Bosnia-Herzegovina, Brazil, Brunei Darussalam, Bulgaria, Canada, Chile, China, Christmas Island, Colombia, Comoros, Costa Rica, Croatia, Cuba, Cyprus, Czech Republic, Denmark, Dominican Republic, Ecuador, El Salvador, Estonia, Ethiopia, Faroe Islands, Fiji, Finland, Former U.S.S.R., Former Yugoslav Republic, France, French Polynesia, Georgia, Germany, Gibraltar, Greece, Guam, Guatemala, Hong Kong, Iceland, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Cocos (Keeling) Islands, Kuwait, Latvia, Lebanon, Lesotho, Lithuania, Luxembourg, Macau, Macedonia, Malaysia, Malta, Mauritius, Mexico, Moldova, Mongolia, Morocco, Nepal, Netherlands, New Caledonia, New Zealand, Nicaragua, Nigeria, Niue, Norway, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russia, Samoa, San Marino, Saudi Arabia, Singapore, Slovakia, Slovenia, South Africa, South Korea, Spain, Sri Lanka, Sweden, Switzerland, Taiwan, Thailand, Tonga, Trinidad and Tobago, Turkey, Ukraine, United Arab Emirates, United

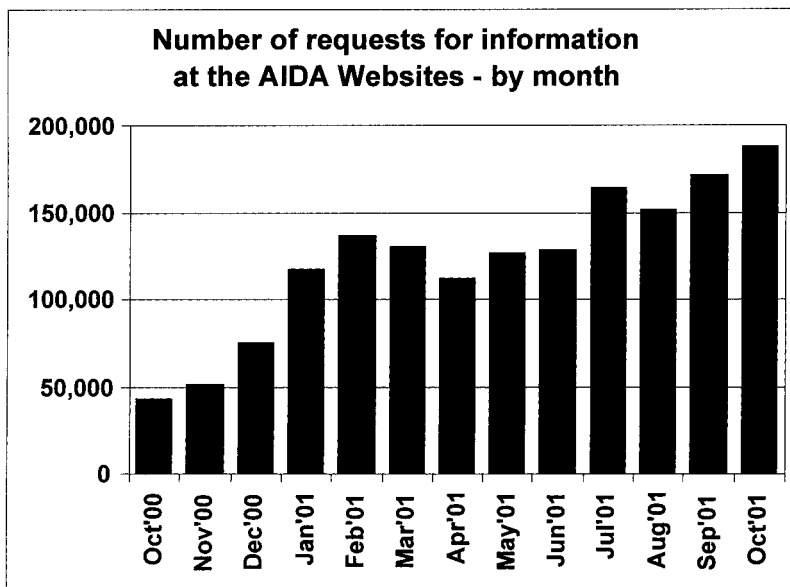


FIG. 3. Number of requests for information (or “hits”) received at the AIDA Website servers each month from October 2000 to October 2001. Raw logstats data from the main AIDA Website (<http://www.2aida.org>) and the first AIDA U.S. mirror site (<http://us.2aida.org>) have been analysed together. (December 2000 data provided from Webalizer online analysis.)

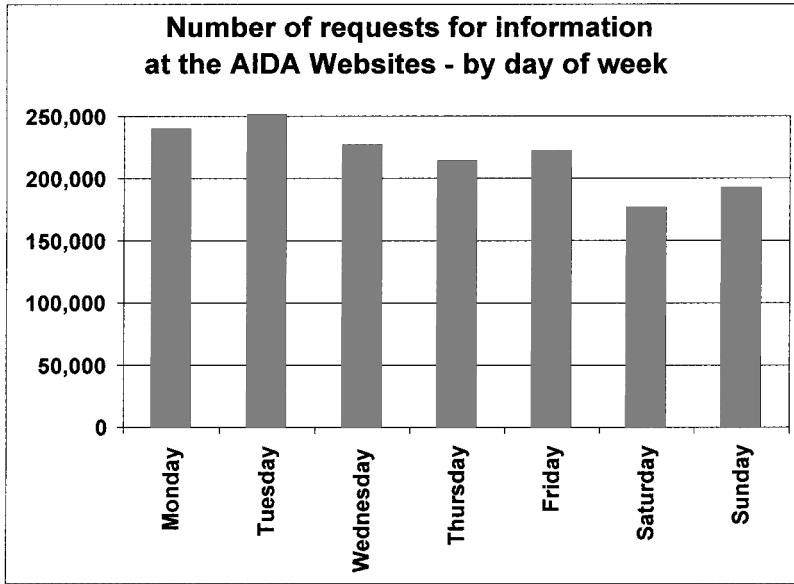


FIG. 4. Overall number of requests for information (or “hits”) received by the AIDA Website servers each day of the week. Slightly lower usage (activity) is seen on the weekend (Saturday and Sunday) compared with during weekdays.

Kingdom, United States, Uruguay, Venezuela, Yemen, and Yugoslavia. The 12 most common countries to access the AIDA Websites during this survey are shown at the bottom right of Table 2.

In terms of actual host servers originating visits it is generally easiest to recognise the

server names of academic or educational establishments, and those of large international companies. In this respect, visits were logged to the AIDA sites (in alphabetical order) from Austria from the Universities of Vienna and Graz, as well as Siemens (Austria).

In Australia visits were logged from the Uni-

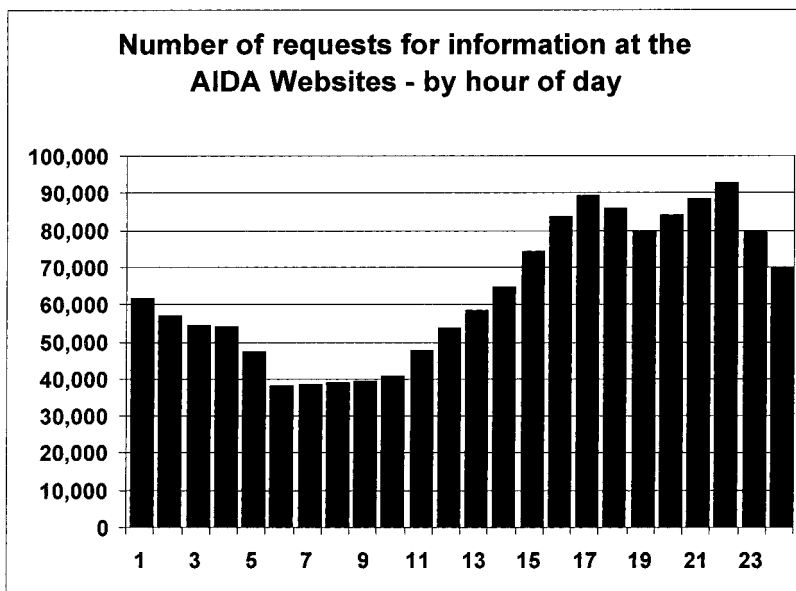


FIG. 5. Overall number of requests for information (or “hits”) received by the AIDA Website servers each hour of the day. All times are given as GMT. Peak usage is at 22:00 h GMT with the lowest activity at 06:00 h GMT.

TABLE 2. COMPARISON OF TECHNICAL DETAILS OF COMPUTER AND OPERATING SYSTEM USAGE, AS WELL AS COUNTRIES ORIGINATING VISITS TO THE AIDA WEBSITES, BETWEEN PREVIOUS AIDA v4 DOWNLOAD AUDITS (VISITOR SELF-REPORTED DATA) AND THE CURRENT LOGSTATS-BASED PILOT SURVEY (AUTOMATICALLY DETECTED DATA)

	<i>AIDA v4.0 download survey</i> <sup>13</sup>	<i>AIDA v4.3 download survey</i> <sup>14</sup>	<i>Current pilot logstats-based Website usage survey</i>
Study period	November 1999–July 2000	July 2000–March 2001	October 2000–November 2001
Data collection type	Downloader self-reported	Downloader self-reported	Automatically detected
32-bit Windows operating system	95.4%	98.2%	89.7%
16-bit Windows operating system	3.2%	0.2%	0.4%
Apple Macintosh computer	1.0%	1.5%	2.2%
Proportion of AIDA downloads due to AIDA v4.3/v4.3a (vs. AIDA v4.1/v4.1a)	Not applicable	98.4%	94.6%
Most common countries originating visits	United States United Kingdom Italy Germany Canada Brazil Spain Poland Australia Netherlands Korea Greece	United States United Kingdom Canada Germany Italy Australia Brazil Netherlands Poland Spain Greece India	United States United Kingdom Canada Australia Italy Germany Netherlands Spain Brazil France Poland Greece

versities of Adelaide, Canberra, Melbourne, and New South Wales, as well as from Flinders Educational Centre, Monash Medical Centre, Newcastle (just north of Sydney), Melbourne Water, Fujitsu (Australia), the Australian Ministry of Defence, the Western Australian Department of Health—and Tasmania, Victoria, and Western Australia government servers.

In Belgium visits were logged from the University of Leuven, and other academic centres.

From Canada visits were logged from London (Ontario), from the Government of British Columbia server, from the Universities of Alberta, Calgary, Waterloo, Guelph, and Toronto, as well as from McGill University, McMaster University, Siemens (Canada), the Canadian Diabetes Association, and the Juvenile Diabetes Research Foundation (Canadian Office).

From Denmark visits were logged from Aalborg University, Bang & Olufsen, Gateway Hospital, and Novo Nordisk (Research).

From France visits were logged from the Universities of Strasbourg and Bordeaux.

From Germany visits were logged from Bayer, Siemens, as well as the Universities of Heidelberg, Leipzig, Dortmund, Dusseldorf, Frankfurt, Hamburg, Hanover, and Mainz.

From Hungary visits were logged from the Semmelweis University of Medicine.

From Italy visits were logged from the University of Trieste.

From New Zealand visits were logged from the Universities of Canterbury and Otago.

From Switzerland visits were logged from the University of Basle and from Disetronic (Switzerland).

From the United Kingdom visits were logged from Bell University, Oxford Brookes University, Coventry University, Birmingham University, Cambridge University, St. George's Hospital Medical School, the United Medical and Dental Schools (of Guy's and St. Thomas' Hospitals), Salford University, Glaxo Wellcome, Bayer (U.K.), British Aerospace Systems, City University, Dundee University, Edinburgh University, Essex University, Glasgow University, Harlow College, Kings College London, North Lincolnshire College, Roehampton University, Strathclyde University, University College London, John Wiley and Sons (publishers), the National Health

Service network, and the British Medical Association.

In the United States visits were logged from Abbott, America Online, Bayer (U.S.A.), Boeing, Compaq, IBM, Johnson & Johnson, the American Diabetes Association, Astra Zeneca, Apple Computers, Federal Express, Merck, Microsoft, Novartis, Pfizer, Eli Lilly, Minimed, Philips, Roche, and Rockwell. Educational establishments that logged visits in the United States include Arizona University, Harvard University, as well as the Universities of Berkeley, Hawaii, Buffalo Medical Center, Indiana, Louisiana, Maryland, Michigan, Oregon, Texas, Virginia, and Wisconsin. Visits were also logged from Columbia University, Cornell University, Dartmouth University, Duke University, Rutgers State University of New Jersey, Emory Medical Center, the Mayo Clinic, Oklahoma State University, University of California at San Francisco, University of California at Los Angeles, the California Institute of Technology, Purdue University, the Medical University of South Carolina, North Carolina State University, the National Aeronautics and Space Administration, the National Institutes of Health, West Texas A&M University, Vanderbilt University, the Food and Drug Administration, Yale Medical Center, as well as from Veterans Medical Centers and military establishments.

This list is far from exhaustive, and focuses mainly on organisations that have readily identifiable host server names. Nevertheless it does offer some indication as to the range of places that have been accessing the AIDA Website servers between October 1, 2000 and November 1, 2001.

Analysis of the logstats files can also provide useful information about which search terms visitors have used to find the Websites via search engines. Table 3 lists these search words, the most common being diabetes (2,518 "hits"). Interestingly the second highest number of search word "hits" (1,602) were for searches on the word "software."

Of the requests received at the AIDA Websites during the study period, 1,196,058 (78.5%) were from visitors using Microsoft's Internet Explorer™ browser, with just 218,460 (14.3%) of the requests being from visitors using the Netscape Navigator™ browser, and just 38,827 (2.5%) be-

TABLE 3. NUMBER OF REQUESTS FOR INFORMATION ("Hits") SORTED BY INDIVIDUAL SEARCH WORDS USED, FOR A SELECTION OF THE 15 MOST COMMON ENTRIES

<i>Search word</i>	<i>"Hits"</i>
diabetes	2,518
software	1,602
glucose	1,370
aida	962
insulin	928
blood	859
freeware	415
diabetic	304
download	253
simulator	234
free	177
program	170
humulin	135
levels	121
action	120

ing from visitors using a Netscape-compatible browser. Only 5,535 "hits" (0.4%) were logged from visitors using the Opera browser, with 2,343 "hits" (0.15%) being recorded from people using WebTV. The remaining "hits" were from unspecified browsers and/or search engine robots spidering the AIDA Websites.

The Analog software was able to resolve operating systems for 1,519,952 of the "hits" to the site. Of these, the vast majority of visitors were using Windows (Microsoft)-enabled computers, with 1,371,710 "hits" (90.2%) being from Windows operating systems [Windows 98, 787,242 "hits" (51.7%); Windows 95, 175,156 "hits" (11.5%); Windows ME, 142,634 "hits" (9.4%); Windows NT, 138,335 "hits" (9.1%); Windows 2000, 113,100 "hits" (7.4%); unknown/32-bit Windows operating systems, 8,898 "hits" (0.6%); 16-bit Windows operating systems, 5,223 (0.3%); Windows 3.1, 1,121 (0.1%)]; the remainder of the Windows "hits" were from unrecognised Windows operating systems.

Just 34,055 "hits" (2.2%) were from Apple Macintosh operating systems, and 10,071 "hits" (0.7%) were from Unix operating systems. As observed in the Web browser section, above, and confirmed in the separate operating system analysis, 2,343 "hits" (0.15%) were from WebTV users. Only 41 "hits" were from BeOS users. The remaining 101,263 "hits" (6.7%) were from unknown operating systems.

The analysis can also provide a summary of the different sort of file types that were downloaded from the AIDA sites during this time. The largest proportion (37.1%) were the executable AIDA software installation files, with 32.1% being accounted for by HyperText Markup Language (HTML) Web pages. Graphics files accounted for 17.0% of the downloads, with portable document format (.pdf) file articles accounting for 4.5% (4,159) "hits."

AIDA online, v2, the Web-based diabetes simulator that operates at [www.2aida.org/online](http://www.2aida.org/online), became operational at the AIDA site in September 2001, and the old v1 AIDA online site was only actually discontinued at the end of December 2001. Therefore during the current study period executable CGI-BIN scripts, such as are used to generate AIDA online's diabetes simulations, only accounted for a relatively small proportion of the usage of the site (30,403 "hits"; 2.75%), with 22,539 AIDA online simulations being run. Overall these simulations constituted just 2% of site traffic (in terms of bandwidth)—mainly because the amount of data that need to be transmitted for an AIDA online simulation is so small (approximately 9 kB).<sup>19,20</sup>

The Analog analysis also permits a review of the Web page URLs (addresses) that are referring visitors to the AIDA Websites. In total over 1,280 referring pages were identified. Interestingly some 2,778 "hits" came via Web-based automated language translation facilities, such as <http://translate.google.com>, <http://www.excite.co.jp>, <http://www.systranlinks.com>, and <http://babelfish.altavista.com>. While these do not offer perfect translations, they can provide the "gist" of what a Web page or Website is about. These facilities involved translation of AIDA Web pages into French, Spanish, German, Italian, Greek, as well as Japanese.

During the period of the current study over 8,050 copies of the AIDA v4 software were downloaded, and over 88,600 separate visits were logged to the AIDA sites. The AIDA Technical Guide—[www.2aida.org/technical](http://www.2aida.org/technical)—was accessed 1,991 times during the survey period. Table 4 summarises the most commonly accessed Web pages and files during the study period. Not surprisingly the home page ([www.2aida.org/aida/intro.htm](http://www.2aida.org/aida/intro.htm)), which may or may

TABLE 4. SELECTION OF MOST COMMONLY ACCESSED PAGES/FILES<sup>a</sup> AT THE AIDA WEBSITES

<i>Web page</i>	<i>"Hits"</i>
intro.htm (homepage)	35,302
simulate.cgi (AIDA online simulation) <sup>b</sup>	22,539
aidadown.htm (main AIDA download page)	10,779
options.htm (AIDA online entry page)	8,622
example.htm (AIDA online fast-track simulation page)	7,060
aidainst.exe (AIDA v4.3 download install file) <sup>c</sup>	6,645
downaida.shtml (AIDA download pop-up link page)	4,004
<a href="http://www.2aida.org/online">www.2aida.org/online</a> (direct URL for AIDA online)	3,844
aidadown2.shtml (first page displayed after visitors reply to the download survey/audit)	3,818
links.htm (AIDA links page)	2,860
caveats.htm (caveats/warnings page)	2,676
aidademo1.htm (first page of AIDA demo)	2,627
download.htm (page with download menu options/HTML links)	2,005
technical.htm (AIDA Technical Guide)	1,991
helpaida.htm (page containing information about helping with AIDA)	1,909
tutorial.htm (first page of Diabetes/Insulin Tutorial) <sup>d</sup>	1,696
manlabel.htm (one graphical summary of AIDA model functions)	1,668
review.htm (user reviews of AIDA v4)	1,367
insulin.htm (details of supported AIDA insulin preparations)	1,302
news.htm (latest AIDA news)	1,235

HTML Web pages can be accessed directly in the [www.2aida.org/aida/directory](http://www.2aida.org/aida/directory).

<sup>a</sup>Does not include frame (menu) bars and subpages called to display frames/pop-ups.

<sup>b</sup>AIDA online only launched at main AIDA Website in September 2001.

<sup>c</sup>Other versions of AIDA v4 logged separated.

<sup>d</sup>Diabetes/Insulin Tutorial launched in September/October 2001.

not be loaded within the Website frames, was the most frequently accessed page. The next most commonly accessed page was the AIDA online Web-based diabetes simulation—a dynamically generated .cgi page—with the third most frequently accessed page being the main AIDA download page ([www.2aida.org/aida/aidadown.htm](http://www.2aida.org/aida/aidadown.htm)) (Table 4).

## DISCUSSION

The process of analysing the raw logstats data files at one level is quite fixed and automated, almost mechanistic. However, some interpretation of the summary findings is really required to draw useful conclusions from the data.

### *Terminology*

Many reports of Website usage, particularly on the Internet, describe the number of "hits" that a site receives. An important thing to appreciate is the relative "unreliability" of "hits" as an index of Website usage. Basically the number of "hits" can somewhat overestimate

activity at a Website, unless it is understood what is actually being counted. As shown in Figure 2, every single request for information from a Website server is recorded in the logstats file. Each one of these entries generates a "hit," which has led some people to describe "hits" more accurately as simply "requests for information."

Let's take an example. Loading a single HTML Web page with no graphics will be logged as a single "hit" in the logstats file. By contrast, loading that same page with one graphic (image) will be logged as two "hits," one for the HTML page and one for the graphic. If that Web page contained five graphics/images, then loading it would be logged as six "hits" (one for the Web page and five for the graphics). It is easy to see how the number of "hits" can be inflated simply by the presence of graphics on a page. Similarly loading each Website frame (menu) bar will increase the number of apparent "hits." As many Web sites/pages use frames and contain dozens of small graphics elements, it will be apparent that the number of "hits" recorded can give rise to a very large number.

If “hits” can be so “unreliable” why are they used? The simple answer is that, as shown in Figure 2, “hits” represent the most basic non-divisible indicator of activity at a site, with all other indices being derived from this. Each “request for information” is logged, so the parameter is useful. It is just the interpretation that is placed on the index that needs to be considered.

You may well ask, are there no alternatives to “hits” or “requests for information” as parameters to monitor? Yes there are. In addition to monitoring the number of “hits,” some logstats analysis software records the number of “visits.” However, this is necessarily an estimate, based on the type of data shown in Figure 2. To assess how many separate visits a site may have received, each set of activity attributable to a single Internet (IP) address may be counted as a single visit. Some other analysis software takes the activity from a particular host before a lull (gap) of at least 30 min—although this is somewhat arbitrary. Another way of estimating activity is by counting the number of unique Internet (IP) addresses to access a particular Website. However, once again, this is dependent on analysis of the same raw logstats data as shown in Figure 2.

Ultimately none of the parameters or indices is perfect. However, the important point is that the number of “hits” recorded by a Website will be much larger than the actual number of Web pages viewed or the number of separate visits logged at the site. Nevertheless it is important to understand what is being monitored to properly interpret the data.

Connected with this, it will be self-evident that the Internet is still a relatively young medium. Therefore it should be pointed out that the analysis and interpretation of Website logstats is not always an exact science. Given this, we believe these indices become of greatest use when followed over time, relative changes in some ways being of more use than absolute values. As such, if a Website is not being updated or revised, then the number of “hits” does offer a useful indicator of increase or decrease in activity at the site over time. Therefore as with much of the data that have been discussed here, having a baseline for comparison is useful.

Furthermore in this current pilot study a particular use of the data has been to compare the observations made with data from previous audits/surveys based on a completely separate methodology, with the intention of confirming the responses in the earlier surveys.<sup>13,14</sup>

#### *Cacheing and under-estimating usage*

One issue that can complicate assessments of Website usage, and that deserves special mention here, is that of cacheing.

There are two major types of cacheing. First, browsers automatically cache files when they are downloaded. This means that if someone visits them again, say the next day, they do not need to download the whole page again. Depending on the settings of their browser, they might check with the server that the page has not changed; in that case, the host server would know about it, and it would be counted as a new request for the page (“hit”). However, the browser might be set up not to check first; then the cached copy of the page would be read again without the host server ever knowing about it.<sup>21</sup>

The other sort of cache is on a larger scale. Almost all Internet service providers (ISPs) now have their own cache. This means that if someone tries to look at a particular page and anyone else from the same ISP has looked at that page recently, the cache will have saved it, and will give it out without ever telling the host server about it. (This applies whatever the browser settings.) So hundreds of people could read a set of Web pages, even though the host server had only sent them out once.<sup>21</sup>

Cacheing may explain, for instance, why when announcements have been sent in the past to the AIDA registration/announcement list—accessible directly at [www.2aida.org/register](http://www.2aida.org/register) (current membership >2,300 people)—the activity seen at the AIDA Website following such an announcement, although increased, does not reflect such a large number of visitors to the site. Obviously it is possible that some people may have not read the announcement, or they may have chosen to ignore it. However, another explanation is that if one person who has an account with a large ISP clicks on the links in the announcement, then the pages will

be cached on the ISP's server, and the next person who requests the same pages—via the same ISP—will be able to access these locally, without requesting them from the AIDA servers. Thereafter it matters little whether two people or 200 people click on links to the same pages, via the same ISP; these requests will not reach the AIDA Website servers, and therefore will not be logged or registered. In this way, the logstats data recorded by a server can substantially underestimate the actual number of people viewing/downloading material from a Website. Given this, the only thing that can be known for certain is the number of requests actually made to the host server, which provides a *minimum* estimate of site usage. More detailed discussion of caching is beyond the scope of this article. However, further information about interpreting Website logstats can be found elsewhere on the Internet (e.g., see Turner,<sup>21</sup> Haigh and Megarity,<sup>22</sup> and Linder<sup>23</sup>).

### *Interpretation*

So what can we learn from this pilot study? Well, the data shown in Figure 4 demonstrate somewhat lower usage of the AIDA Website over the weekend. This may be put down to people having other things to do on Saturdays and Sundays. Also, this difference may in part reflect that some people are visiting the Websites using computer/Internet access from their place of work, and this access may be reduced at weekends. This interpretation is reinforced by the wide range of companies that are identified as the owners of source computers responsible for visits to the AIDA sites. With the exception of some pharmaceutical companies with specific diabetes interests, it is presumed that the majority of these visits are being made by people with a personal interest in diabetes—who just happen to work for these companies.

As shown in Figure 5, there is a consistent nadir in usage at the site at 6 a.m.—so this has been the time selected for automated house-keeping activities at the Websites, as these will have the minimum impact on actual site visitors.

Similarly, one of the principal uses for this pilot study has been to validate certain data that we have collected in earlier studies<sup>13,14</sup> via

other means. Table 2 shows a comparison of data collected in two previous studies<sup>13,14</sup> compared with similar data from the current pilot study. In the two earlier studies conducted from November 1999 to July 2000<sup>13</sup> and July 2000 to March 2001<sup>14</sup> people downloading the AIDA software were asked various questions including technical details about which operating system they were running, what sort of computer they were using, what version of the AIDA software they were downloading, as well as which country they were from. The feedback is based on self-reported data. Therefore it is interesting to compare this with the results obtained from the current pilot study based on automatically extracted data from the AIDA logstats files. Usage of old 16-bit Windows operating systems will obviously decrease over time. Given this, allowing for the different time periods, as well as the caveats highlighted above about how the Website server logstats data are collected and should be interpreted, the results do seem rather consistent irrespective of the method of data collection (Table 2).

In this respect the data that we are obtaining from the logstats files do appear to help confirm the responses we have previously obtained from downloaders at the Website. This validation of self-reported data previously obtained by other means is important because it lends credence to some of the other self-reported answers given by AIDA downloaders (Fig. 1). Furthermore it suggests that earlier reports<sup>13,14</sup> of over half of AIDA downloaders being people with diabetes, with 12–13% being relatives of patients, and another 12–13% being doctors (Table 1), may well be accurate.

Other interesting observations to come out of this pilot study include the use by visitors of automated Web-based translation facilities to translate pages from the AIDA Websites into other languages. There is a growing realisation that a large proportion of Internet users do not speak (or read) English. Given this, there has been interest expressed in translating AIDA/AIDA online and/or portions of the AIDA Website into other non-English languages. Some of the issues surrounding such translation work are discussed in detail at [www.2aida.org/translation](http://www.2aida.org/translation)

Furthermore, in collaboration with other non-commercial Websites, Diabetes/Insulin Tutorial pages from [www.2aida.org/tutorial](http://www.2aida.org/tutorial) have already begun being translated into Russian and Italian (see [www.2aida.org/russian](http://www.2aida.org/russian) and [www.2aida.org/italian](http://www.2aida.org/italian)). It is hoped over time to see this material translated into other languages—if other non-English language diabetes Websites can be found that are interested in such collaborations.

## CONCLUSIONS

This study has yielded an interesting insight into how the AIDA Website is being used, and appears to confirm data collected previously by other means.

As has been discussed above, it is important to appreciate that numbers of “hits,” which give large numbers, do not equal number of site visitors or Web pages viewed; the latter two indices will be much smaller. Also, potentially, because of local cacheing between the visitor’s computer and the AIDA servers, it is possible that many more people may have viewed the AIDA Web pages and downloaded copies of the software than is necessarily made apparent by the data in the logstats files. All we can say for sure is that *at least* the numbers of requests for information recorded in the raw logstats data files have been received by the AIDA servers. While this is not an exact science it does, we believe, provide useful information—especially when changes are monitored over time. Furthermore, while the absolute number of “hits”/visitors may be an underestimate because of cacheing, it might be expected that the proportions or percentages should be reliable. In this respect, as shown in Table 2, remarkably consistent observations have been made over a substantial period of time using different study methodologies—offering some useful confirmation of earlier survey/audit methods.

Finally, while Webalizer is useful for online analysis and summarising the results on a short-term/day-by-day, ongoing basis, Analog allows a much more detailed assessment of the data, and therefore is planned to be used in future studies. Connected with this, to simplify

comparisons between different methods of data collection, it is planned in future to standardise the time period for such audits so that there is a standard start and end time for these surveys. Probably, in future, we will aim to run analyses from the beginning of March, on an annual or bi-annual basis.

## FURTHER TOPICS

If you would like to suggest further topics or Websites for future “Diabetes Information Technology & WebWatch” columns, please e-mail information—with a brief description of the site/suggestion—to Dr. E.D. Lehmann: [info-www@2aida.org](mailto:info-www@2aida.org) (please write Diabetes WebWatch in the subject line). You can also fax information to: (503) 218-0828, quoting Diabetes Information Technology & WebWatch.

## REFERENCES

1. Lehmann ED, Deutsch T, Broad D: AIDA: An Educational Simulator for Insulin Dosage and Dietary Adjustment in Diabetes. London: British Diabetic Association, 1997.
2. Lehmann ED: Experience with the Internet release of AIDA v4—an interactive educational diabetes simulator. *Diabetes Technol Ther* 1999;1:41–54.
3. Lehmann ED: User experience with the AIDA interactive educational virtual diabetes patient simulator. *Diabetes Technol Ther* 2000;2:165–171.
4. Lehmann ED, Tatti P: Questionnaires for a randomized controlled trial methodology to evaluate the teaching utility of diabetes simulation programs. *Diabetes Technol Ther* 2001;3:293–305.
5. Tatti P, Lehmann ED: Utility of the AIDA diabetes simulator as an interactive educational teaching tool for general practitioners (primary care physicians). *Diabetes Technol Ther* 2001;3:133–140.
6. Tatti P, Lehmann ED: Use of the AIDA diabetes simulation software—[www.2aida.org](http://www.2aida.org)—as an interactive educational tool for teaching student nurses. *Diabetes Technol Ther* 2001;3:655–664.
7. Lehmann ED: Further user comments regarding usage of an interactive educational diabetes simulator (AIDA). *Diabetes Technol Ther* 2002;4:121–135.
8. Lehmann ED, Tatti P: Randomised controlled trial design for simulator use in diabetes education: some issues for consideration [letter]. *Diabetes Technol Ther* 2002;4:258–269.
9. Tatti P, Lehmann ED: Using the AIDA—[www.2aida.org](http://www.2aida.org)—diabetes simulator. Part 1: Recommended

- guidelines for health-carers planning to teach with the software. *Diabetes Technol Ther* 2002;4:401-414.
10. Lehmann ED, Tatti P: Using the AIDA—[www.2aida.org](http://www.2aida.org)—diabetes simulator. Part 2: Recommended training requirements for health-carers planning to teach with the software. *Diabetes Technol Ther* 2002;4:717-732.
  11. Lehmann ED, Deutsch T: A physiological model of glucose-insulin interaction in type I diabetes mellitus. *J Biomed Eng* 1992;14:235-242.
  12. Lehmann ED: The freeware AIDA interactive educational diabetes simulator—<http://www.2aida.org>—(2) Simulating glycosylated haemoglobin (HbA<sub>1c</sub>) levels in AIDA v4.3. *Med Sci Monit* 2001;7:516-525.
  13. Lehmann ED: The freeware AIDA interactive educational diabetes simulator—<http://www.2aida.org>—(1) A download survey for AIDA v4.0. *Med Sci Monit* 2001;7:504-515.
  14. Lehmann ED: Who is downloading the freeware AIDA v4.3 interactive educational diabetes simulator? An audit of 2,437 downloads. *Diabetes Technol Ther* 2002;4:467-477.
  15. Lehmann ED: Why are people downloading the freeware AIDA diabetes computing software program: a pilot study. *Diabetes Technol Ther* 2002;4:793-808.
  16. Lehmann ED: Preliminary experience with the Internet release of AIDA—an interactive educational diabetes simulator. *Comput Methods Programs Biomed* 1998;56:109-132.
  17. Lehmann ED: Spontaneous comments from users of the AIDA interactive educational diabetes simulator. *Diabetes Educ* 2000;26:633-643.
  18. Lehmann ED: Short user comments ('sound bites') regarding usage of AIDA v4—<http://www.2aida.org>—an interactive educational diabetes simulator. *Diabetes Technol Ther* 2000;2:663-666.
  19. Lehmann ED, DeWolf DK, Novotny CA, Gotwals RR, Rohrbach RP, Blanchard SM: Dynamic interactive educational diabetes simulations using the World Wide Web [abstract]. *Diabetes Technol Ther* 2001;3:A17-A20.
  20. Lehmann ED: Usage of a diabetes simulation system for education via the internet [letter]. *Int J Med Inform* 2003;69:63-69.
  21. Turner S: How the web works. Available from <http://www.analog.cx/webworks.html>. Last updated June 25, 2002. Accessed November 3, 2002.
  22. Haigh S, Megarity J: Measuring web site usage: log file analysis. Available from <http://www.nlc-bnc.ca/publications/1/p1-256-e.html>. Last updated February 27, 2001. Accessed November 6, 2002.
  23. Linder D: Interpreting WWW statistics. Available from <http://www.arjo.ch/etc/webstats.html>. Last updated January 1994. Accessed November 3, 2002.

Address reprint requests to:

*Dr. Eldon D. Lehmann  
c/o AIDA Diabetes Simulator  
Development Team  
P.O. Box 46104  
London, EC2Y 8WN, UK*

*E-mail:* [info-www@2aida.org](mailto:info-www@2aida.org)

*Web:* <http://www.2aida.org/lehmann>