

Aktuariell diskontiertes Überleben aus webbasierten Rohdatenbanken

Alexander Wagner, Karl J. Krobot
MSD SHARP & DOHME GMBH
Lindenplatz 1
85540 Haar bei München
alexander_wagner@msd.de

Zusammenfassung

Hintergrund und Zielsetzung: In der demographischen, gesellschaftspolitischen und gesundheitsökonomischen Forschung nimmt die vergangene, aktuelle oder zukünftige Mortalität eine zentrale Rolle ein. Für viele Fragestellungen ist Lebenserwartung krankheitsfrei zu schätzen, bereits auf der Ebene der Sterbetafel zu diskontieren [1] oder mit Konfidenzintervallen zu versehen [2]. Diese Berechnungen sind komplex, so dass wir ein System entwickelten, welches auf webbasierte Rohdaten zugreift, diese einliest, zusammenführt, analysiert und die Ergebnisse vollautomatisch in Echtzeit ausgibt. Wir präsentieren das System am Beispiel schlaganfallfreien Überlebens basierend auf online-Datenbanken des Robert-Koch-Instituts [3] und des Statistischen Bundesamtes [4].

Schlüsselwörter: Sterbetafel, SAS, AutoIt, MS-PowerPoint, OWC11.SpreadSheet, OWC11.ChartSpace

1 Material und Methoden

Das System wurde mit SAS 9.2 [5], AutoIt 3.3.0 [6], MS-Office Web Komponenten [7,8] unter Windows XP realisiert. AutoIt greift auf die webbasierten Rohdaten zu und liest diese ein. Temporäre Dateien (in xls-, csv-, html-, txt-Format) werden mit Hilfe von SAS integriert und analysiert. Die Präsentation der SAS-Ausgabeergebnisse in MS-PowerPoint erfolgt sodann dynamisch und multifunktional. Für diesen Zweck verwenden wir die MS-Office Web Komponenten OWC11.SpreadSheet und OWC11.GraphSpace [7].

2 Systemkomponenten und Arbeitsphasen

Informationsressourcen des Systems sind webbasierte Datenbanken der Bundesgesundheitsberichtserstattung [3] und des statistischen Bundesamtes (DESTATIS) [4]. Methodengrundlage sind Sterbetafeln [2] sowie das aktuarielle Diskontieren [1].

Die drei Phasen des Systems werden nachfolgend besprochen.

3 Phase I. Datengewinnung aus den Internetquellen und Aufbau der Datenbank

Die Datengewinnung aus dem Internet wird mit Hilfe von AutoIt realisiert. Die Datenbank wird unter SAS aufgebaut und gepflegt (Abb. 1).

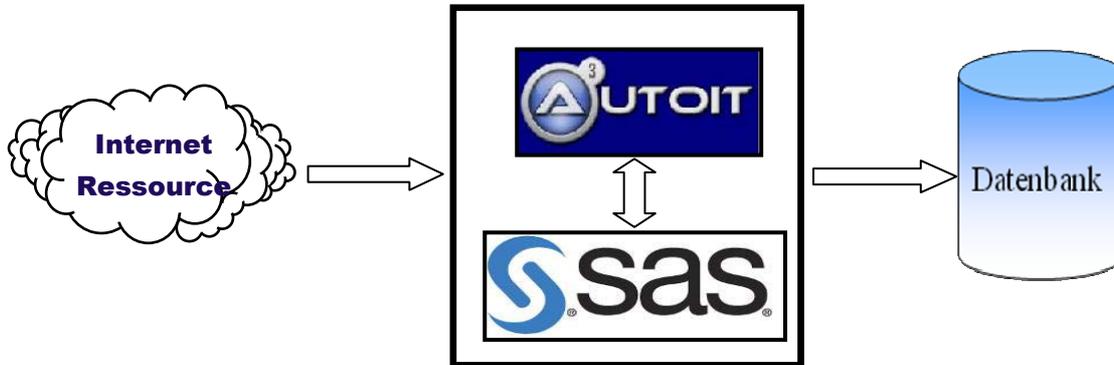


Abbildung 1: Schema der Datengewinnung und der Datenbankaufbau

Unten findet man ein Fragment des Codes des Autoit-Programms. Dieses Programm baut einen Datenzugriff zur webbasierten Rohdatenbank der Gesundheitsberichtserstattung des Bundes auf.

```
#include <GUIConstants.au3>
#include <GUIConstantsEx.au3>
#include <WindowsConstants.au3>
#include <StaticConstants.au3>
#include <IE.au3>
Navi()
While 1
WEnd
Func Navi()
$oIE = ObjCreate("Shell.Explorer.2")
$oIE.navigate ("about:blank")
_IENavigate($oIE, "http://www.gbe-bund.de")
_IELinkClickByText($oIE, "Gesundheitliche Lage")
SLEEP(500)
_IELinkClickByIndex ($oIE, 38)
SLEEP(500)
_IELinkClickByIndex ($oIE, 38)
SLEEP(1500)
_IELinkClickByIndex ($oIE, 84)
SLEEP(1500)
_IELinkClickByIndex ($oIE, 50)
SLEEP(2500)
_IELinkClickByIndex ($oIE, 51)
SLEEP(2500)
```

```

_IELinkClickByIndex ($oIE, 52)
SLEEP(2500)
_IENavigate($oIE, "http://www.gbe-bund.de/oowa921-
install/servlet/oowa/aw92/WS0100/_XWD_PROC?_XWD_2/1/XWD_CUBE.DRILL/_
XWD_32/D.000/3722")
SLEEP(2500)
_IENavigate($oIE, "http://www.gbe-bund.de/oowa921-
install/servlet/oowa/aw92/WS0100/_XWD_PROC?_XWD_2/1/xs_set_xs_werte_
jn/D.000/yes")
SLEEP(2500)
.....
.....
Mehere AutoIt-Programmzeilen
.....
.....
EndFunc

```

Abbildung 2 stellt das Hauptmenü des Informationssystems der Gesundheitsberichtserstattung des Bundes dar. Abb. 3 gibt die Ad-hoc-Tabelle „Sterbefälle je 100 Einwohner (ab 1988)“ der webbasierten Datenbank wieder. Der AutoIt-Programmcode des Zugriffes zu dieser Tabelle lautet:

```

Func INI_SAVE($K)
    SEND ("^a")
    SLEEP(500)
    SEND ("^a")
    SLEEP(500)
    SEND ("^c")
    SLEEP(500)
    SEND ("^c")
    SLEEP(500)
    $X = ClipGet()
    Sleep( 500)
    Send ("{ENTER}")
    $Text = StringSplit($X, @CRLF)
    For $II = 2 To $Text[0]-1
        If ($Text[$II] <> " ") Then
            IniWrite (@ScriptDir & "\GBE.ini", $K, $II, $Text[$II])
        EndIf
    Next
EndFunc

```



Abbildung 2: Hauptmenü des Informationssystems der Gesundheitsberichterstattung des Bundes

Sterbefälle je 100.000 Einwohner (ab 1998). Gliederungsmerkmale: Jahre, Region, Alter, Geschlecht, ICD-10
 Die Tabelle bezieht sich auf:
 Alter: 60 bis unter 65 Jahre, Geschlecht: Männlich, Region: Deutschland

ICD10	Jahr (Jahre absteigend)									
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
A00-T98 Alle Krankheiten und Folgen äußerer Ursachen	1564,1	1503,9	1454,6	1403,6	1379,1	1365,0	1320,7	1298,6	1262,5	1257,3
A00-E99 Bestimmte infektiöse und parasitäre Krankheiten	16,1	17,5	18,1	20,0	18,3	16,7	17,2	18,5	16,6	18,7
C00-D48 Neubildungen	592,0	504,3	509,0	520,0	551,7	542,8	540,3	527,2	522,9	514,0
D50-D90 Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems	1,8	2,6	2,7	2,3	2,2	2,9	2,5	2,6	2,2	2,2
E00-E90 Endokrine, Ernährungs- und Stoffwechselkrankheiten	33,6	32,7	30,9	30,0	30,2	32,8	32,2	34,3	30,8	32,7
F00-F99 Psychische und Verhaltensstörungen	27,8	27,9	26,7	22,9	24,3	23,6	23,1	23,9	25,2	24,5
G00-H95 Krankheiten des Nervensystems und der Sinnesorgane	20,1	19,4	20,1	22,4	22,9	21,4	22,3	22,2	24,1	22,8
I00-I99 Krankheiten des Kreislaufsystems	561,8	531,8	486,0	463,8	444,5	435,1	405,1	391,1	373,2	369,6
I00-I02 Akutes rheumatisches Fieber	-	0,1	0,0	0,1	0,1	-	0,0	0,0	-	-
I05-I09 Chronische rheumatische Herzkrankheiten	3,0	2,9	2,6	2,2	2,1	2,1	1,7	1,9	1,7	1,6
I10-I15 Hypertonie (Hochdruckkrankheit)	15,1	15,9	15,0	16,9	17,2	16,9	18,6	16,6	17,2	16,7
I20-I25 Ischämische Herzkrankheiten	317,2	300,0	268,3	253,4	237,5	234,5	216,9	208,3	197,9	192,0
I21-I22 Herzinfarkt	210,9	195,6	173,9	163,5	148,4	145,9	134,8	132,3	124,2	117,8
I26-I28 Pulmonale Herzkrankheit und Krankheiten des Lungenkreislaufes	16,7	16,3	15,2	14,2	13,4	12,7	13,3	13,0	13,1	13,8
I30-I52 Sonstige Formen der Herzkrankheit	93,5	91,5	89,2	87,4	85,8	85,0	75,0	73,8	66,5	73,4
I60-I69 Zerebrovaskuläre Krankheiten	81,0	72,9	67,5	61,3	60,0	55,5	52,0	50,4	49,9	45,4
I60-I61 Subarachnoidalblutung und intrazerebrale Blutung	25,0	21,6	21,3	19,2	18,7	17,6	15,6	15,7	14,9	14,8
I70-I79 Krankheiten der Arterien, Arteriolen und Kapillaren	27,3	25,1	22,2	21,9	22,5	23,1	22,1	21,6	22,2	22,3
I80-I89 Krankheiten der Venen, der Lymphgefäße und der Lymphknoten	7,5	7,0	6,0	6,4	5,8	5,2	5,4	5,2	4,5	4,3
I85-I89 Sonstige und nicht näher bezeichnete Krankheiten des Kreislaufsystems	0,3	0,1	0,1	0,1	0,1	-	0,1	0,1	0,1	-
J00-J99 Krankheiten des Atmungssystems	70,8	71,8	69,0	63,0	66,6	69,0	63,4	69,5	63,7	71,6
K00-K93 Krankheiten des Verdauungssystems	118,1	115,3	109,6	108,4	104,5	102,7	97,6	96,1	91,8	89,5
L00-L99 Krankheiten der Haut und der Unterhaut	0,4	0,3	0,4	0,2	0,1	0,5	0,4	0,5	0,4	0,5
M00-M99 Krankheiten des Muskel-Skelett-Systems und des Bindegewebes	2,4	2,6	2,0	2,3	2,7	2,8	2,3	2,3	2,9	2,5
N00-N99 Krankheiten des Urogenitalsystems	13,9	12,9	11,0	11,3	9,7	11,0	9,4	10,5	10,0	10,5
O00-O99 Schwangerschaft, Geburt und Wochenbett	-	-	-	-	-	-	-	-	-	-
P00-P96 Bestimmte Zustände, die ihren Ursprung in der Perinatalperiode haben	-	-	-	-	-	-	0,0	0,1	0,0	0,0
Q00-Q99 Angeborene Fehlbildungen, Deformitäten und Chromosomenanomalien	0,8	0,6	0,8	1,1	1,2	1,2	1,6	1,4	1,5	1,2
R00-R99 Symptome und abnorme klinische und Laborbefunde, die demorts nicht klassifiziert sind	42,6	43,8	47,2	45,5	42,8	43,7	43,8	42,7	41,2	47,5
S00-T98 Verletzungen, Vergiftungen und bestimmte andere Folgen äußerer Ursachen	61,3	60,4	60,5	59,8	57,3	58,8	59,2	55,8	56,0	49,5

Abbildung 3: Ad-hoc-Tabelle „Sterbefälle je 100.000 Einwohner (ab 1988)“

In Abbildung 4 sieht man das Datenfragment des GBE.ini Files und unten den Programmcode, mit dessen Hilfe die Datei erstellt wurde (Func INI_SAVE).

```

GBE.ini - Notepad
File Edit Format View Help
58=
59=Die Tabelle bezieht sich auf:
60=
61=Alter: 1 Jahr bis unter 5 Jahre, Geschlecht: Männlich, Region: Deutschland
62=
63=
64=
65=Aktion zurückgesetzt! Informationen zur Handhabung der Ad-hoc-Tabellen.
66=
67=
68=
69=ICD10 Jahr (Jahre absteigend)
70=
71=1998 1999 2000 2001 2002 2003 2004 2005 2006 2007
72=
73=A00-T98 Alle Krankheiten und Folgen äußerer Ursachen 28,6 28,8 26,8 28,0 22,5 24,3 21,5 22,6 20,6 2
74=
75= A00-B99 Bestimmte infektiöse und parasitäre Krankheiten 1,4 1,9 1,3 1,7 1,1 1,4 1,0 1,5 1,8 1,1
76=
77= C00-D48 Neubildungen 2,4 4,0 3,4 4,0 2,5 2,6 3,2 3,3 3,1 3,7
78=
79= D50-D90 Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte
80=
81= Störungen mit Beteiligung des Immunsystems 0,1 0,5 0,5 0,4 0,6 0,1 0,2 0,3 0,3 0,1
82=
83= E00-E90 Endokrine, Ernährungs- und Stoffwechselkrankheiten 0,9 0,7 1,4 0,9 0,7 1,1 1,3 0,5 1,1 1,
84=
85= F00-F99 Psychische und Verhaltensstörungen 0,1 - 0,1 - - - - - 0,1
86=
87= G00-H95 Krankheiten des Nervensystems und der Sinnesorgane 2,7 1,8 2,1 2,2 1,6 2,6 2,2 2,3 1,3 2,
88=
89= I00-I99 Krankheiten des Kreislaufsystems 2,1 2,1 1,6 1,4 1,4 1,7 1,4 1,1 1,3 1,3
90=
91= I00-I02 Akutes rheumatisches Fieber - - - - - - - - - -
92=
Ln 1, Col 1

```

Abbildung 4: Datenfragment der GBE.ini Datei

In Abbildung 5 ist ein Datenfragment des SAS-Datasets GBE_AUTO zu sehen, sowie weiter unten der SAS-Programmcode, mit dessen Hilfe diese Datasets aus der Ini-Datei erstellt wurden.

```

Data GBE.GBE_Auto(KEEP= ALTER SEX ICD Diag J1998 J1999 J2000 J2001
J2002 J2003 J2004 J2005 J2006 J2007);
    FORMAT Num 8. ALTER $24. SEX $9. ICD $7. Diag $200. J1998 J1999
J2000 J2001 J2002 J2003
    J2004 J2005 J2006 8.;
    ARRAY YEAR[10] J1998 J1999 J2000 J2001 J2002 J2003 J2004 J2005
J2006 J2007;
    LENGTH RICD $32767;
    infile "C:\KSFE2009\Programme\GBE.TXT" delimiter='09'x
        MISSOVER DSD lrecl=32767 firstobs=2 ;
    informat VAR $32767. ;
    input VAR $;
    If Substr(VAR, 1, 1) = "[" Then Num = 0; Else
    Num = LEFT(SCAN(VAR, 1, "="))+0;
    VAR = LEFT(SCAN(VAR, 2, "="));
    IF (NUM=63 OR (91 <= NUM <= 117)) and VAR > " " THEN DO;
        IF NUM = 63 THEN DO;

```

	ALTER	SEX	ICD	Diag	J199	J1999	J2000
327	60 bis unter 65 Jahre	Männlich	I00-I02	Akutes rheumatisches Fieber	.	0	0
328	60 bis unter 65 Jahre	Männlich	I05-I09	Chronische rheumatische Herzkrankheiten	3	3	3
329	60 bis unter 65 Jahre	Männlich	I10-I15	Hypertonie (Hochdruckkrankheit)	15	16	15
330	60 bis unter 65 Jahre	Männlich	I20-I25	Ischämische Herzkrankheiten	317	300	268
331	60 bis unter 65 Jahre	Männlich	I21-I22	Herzinfarkt	211	196	174
332	60 bis unter 65 Jahre	Männlich	I26-I28	Pulmonale Herzkrankheit und Krankheiten des Lungenkreislaufes	17	16	15
333	60 bis unter 65 Jahre	Männlich	I30-I52	Sonstige Formen der Herzkrankheit	94	92	89
334	60 bis unter 65 Jahre	Männlich	I60-I69	Zerebrovaskuläre Krankheiten	81	73	68
335	60 bis unter 65 Jahre	Männlich	I60-I61	Subarachnoidalblutung und intrazerebrale Blutung	25	22	21
336	60 bis unter 65 Jahre	Männlich	I70-I79	Krankheiten der Arterien, Arterien	27	25	22

Abbildung 5: Datenfragment des SAS-Datasets GBE

```

ALTER=LEFT(SCAN(SUBSTR(VAR, INDEX(VAR, ":")+1), 1, ","));
IF Substr(ALTER, 1, 1) = "U" Then ALTERN = 0; Else
ALTERN = Substr(ALTER, 1, 2)+0;
SEX    = LEFT(SCAN(SCAN(VAR, -2, ":"), 1, ","));
RETAIN ALTER SEX ALTERN;

END;
ELSE DO;
Len = -1;
IF NUM ne 113 THEN Link Z;
ICD = SCAN(VAR, 1, " ");
Diag = SUBSTR(VAR, Length(ICD)+1);
Diag = SUBSTR(Diag, 1, Length(Diag)-Len);
IF Num not in (113, 115) then OUTPUT;
    IF Num = 113 then DO;
        RICD = ICD;
        RDiag = Trim(Diag);
        retain RICD RDIAG;
        END;
        IF Num = 115 then DO;
            ICD = RICD;
            Diag = Trim(RDiag)||Trim(ICD)||Trim(Diag);
            OUTPUT;
        END;
    END;
END;
```

```

        END;
    END;
RETURN;
Z:
DO I = 1 TO 10;
    YEAR[11-I]=TRANSLATE (TRANSLATE (SCAN (VAR, -I, ""), ".", ", "), ".", "-")+0;
    Len + Length(SCAN (VAR, -I, " ")) +1;
END;
RETURN;
RUN;

```

Abbildung 6 stellt die DESTATIS Internet-Seite dar.

Unten ist ein Fragment des AutoIt-Programms wiedergegeben, mit dessen Hilfe der Zugriff zur webbasierten Rohdatenbank DESTATIS und die Speicherung der Excel-Tabelle „Sterbetafel, männlich“ realisiert wird.

```

$oIE = ObjCreate("Shell.Explorer.2")
$oIE.navigate ("about:blank")
_IENavigate($oIE, "http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/Navigationsknoten_Startseite1.psml")
_IELoadWait ($oIE)
_IENavigate($oIE,
"http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/Statistiken/Bevoelkerung/GeburtenSterbefaelle/Tabellen,templateId=renderPrint.psml__nnn=true")
_IELoadWait ($oIE)
    SEND("{F10}")
    SLEEP(1000)
    SEND("f")
    SLEEP(500)
    SEND("a")
    SLEEP(500)
    SEND(@ScripDir & "\GERMANY")
.....
Mehrere AutoIt Programmzeilen
.....
_IEQuit ($oIE)

```

The screenshot shows the main menu of the German Federal Statistical Office (Destatis). At the top, there is a navigation bar with the logo 'STATIS wissen.nutzen.' and a search bar. Below the navigation bar, there is a breadcrumb trail: 'Sie befinden sich hier: Startseite'. The main content area is divided into several sections. On the left, there is a sidebar menu with categories like 'Indikatoren', 'Themen', 'Services', and 'Datenbanken'. The main content area displays several news items, including 'Januar bis November 2008: Deutsche Ausfuhren nach China um 14,3% gestiegen' and 'Erwerbsbeteiligung älterer Arbeitnehmer'. On the right, there is another sidebar with a list of services and research areas.

Abbildung 6: DESTATIS-Hauptmenu

Phase II. Datenanalyse u. Vorbereitung der Ausgabeergebnisse

Die Datenanalyse und die Entwicklung der tabellarischen und graphischen Ausgabeergebnisse führen wir mit Hilfe von SAS durch. Das Schema des Entwicklungsprozesses zeigt Abbildung 7. Endergebnis der Datenanalyse sind die Tabellen: Sterbetafel (männlich, weiblich), Schlaganfallfreie Sterbetafel (männlich, weiblich), diskontierte Sterbetafel (männlich, weiblich), Lebenserwartung (männlich, weiblich), Schlaganfallfreie Lebenserwartung (männlich, weiblich) und diskontierte Lebenserwartung (männlich, weiblich).

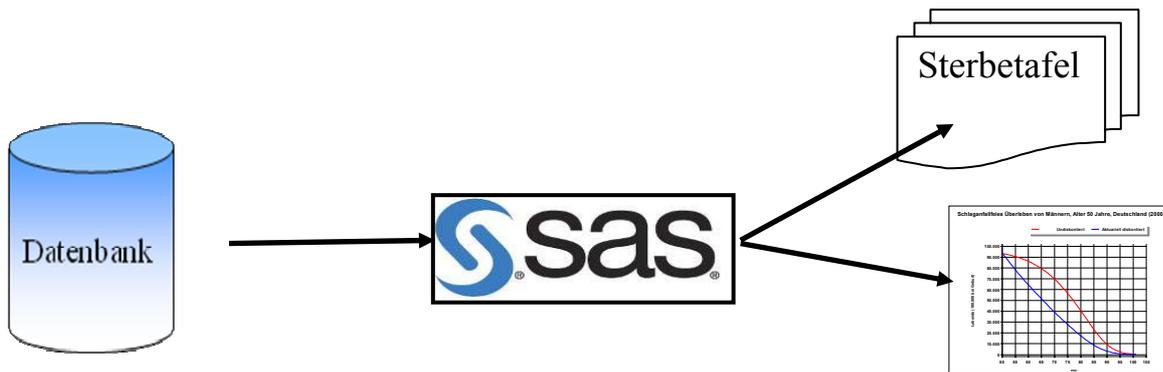


Abbildung 7: Schema der Ausgabeergebnisseentwicklung

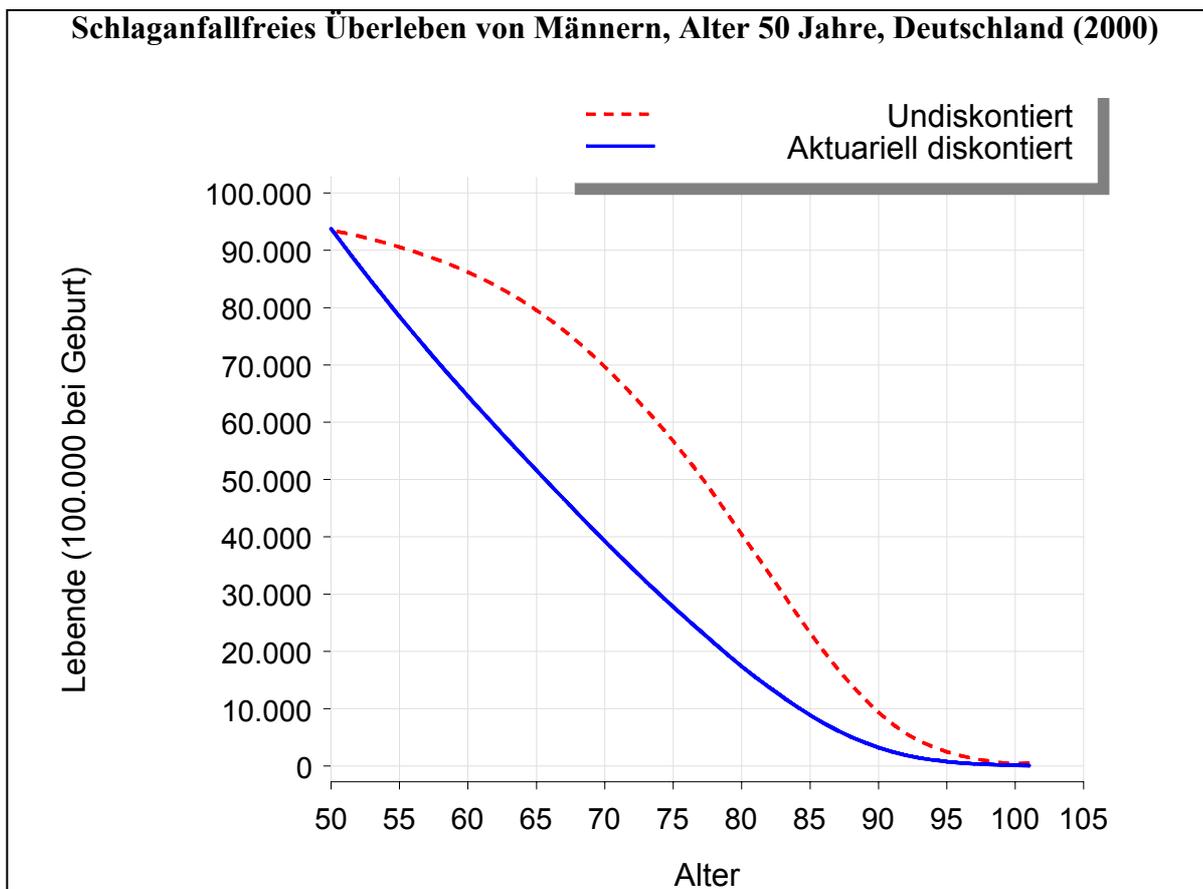


Abbildung 8: Undiskontierte/diskontierte Überlebensgraphik

Phase III. Visualisierung u. Darstellung der Ausgabeergebnisse

Die Visualisierung besteht aus der Konvertierung der SAS-Datasets in OWC11.SpreadSheet, Erzeugung der tabellarisch-graphischen Objekte und ihrer dynamischen Präsentation.

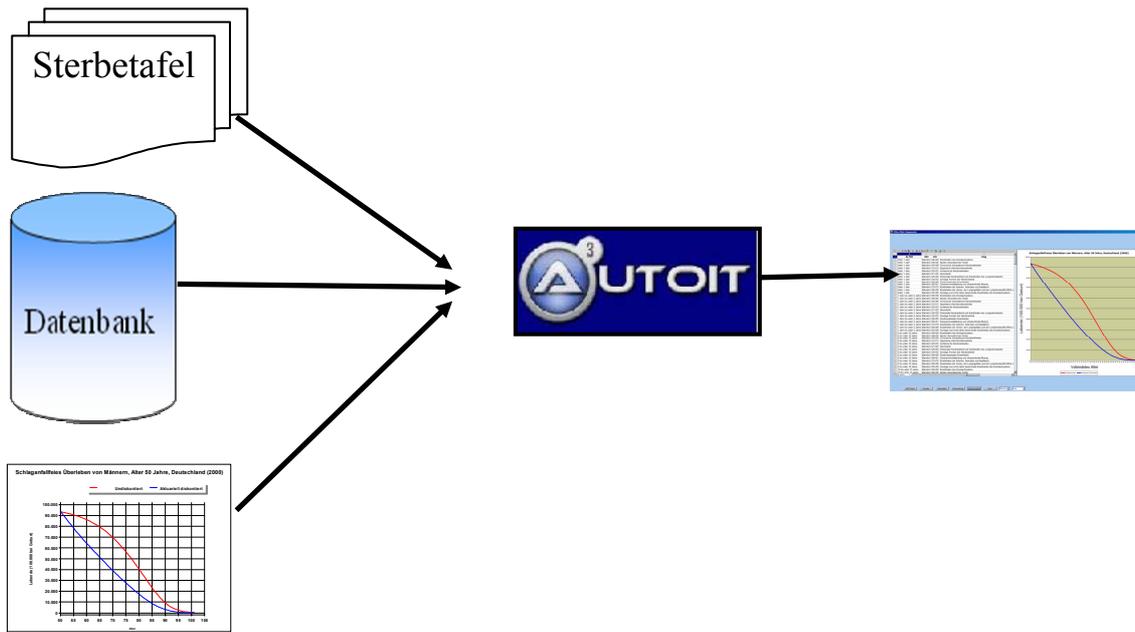


Abbildung 9: Das Visualisierungsschema der Datenbank und der Ausgabeergebnisse

In Abbildung 10 ist ein Visualisierungsbeispiel zu sehen.

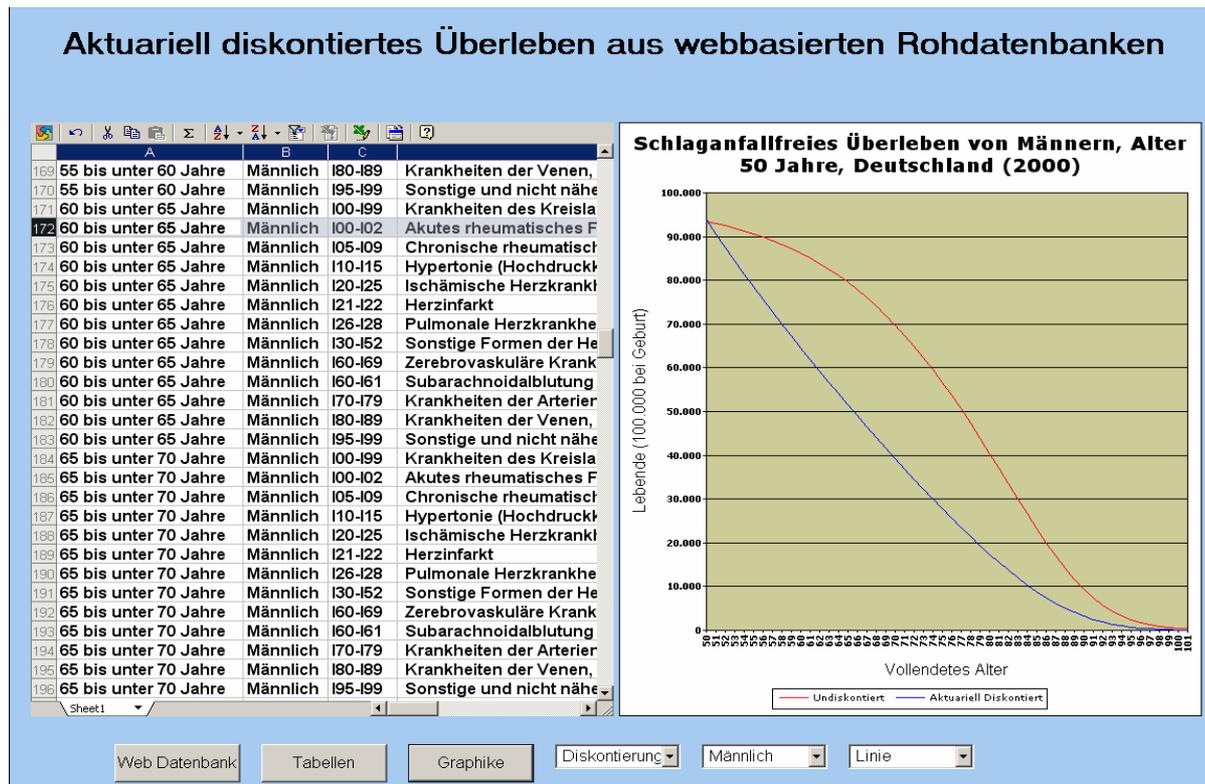


Abbildung 10: Visualisierungsbeispiel einer Tabelle und einer Graphik

Auf der Begleit-CD finden Sie die AutoIT-Programmblöcke GBE_Table, der die SAS-Datasets in OWC11.spreadsheet visualisiert, und FunctionGraph3 zum Aufbau der Graphiken, die anschließend mit OWC11.ChartSpace präsentiert werden.

4 Ergebnisse

Vor der Fertigstellung der Technologie nahm die Bearbeitung einer Fragestellung zur Sterbestatistik (www.gbe-bund.de) Tage in Anspruch. Das nunmehr vollautomatisierte System sorgt für einen strukturierten Datenaufbau und steuert den Prozess "Internet-Analyse-Präsentation" innerhalb weniger Minuten. Die Abbildung zeigt aktuarielles Diskontieren. Die korrekt diskontierte Lebenserwartung leitet sich aus der Fläche unter der linkskonvexen Kurve ab und ist erheblich kleiner als nach dem weitverbreiteten Ansatz $\sum_{n=0}^{\infty} 1/(1+0,05)^n$, $n=0 \dots LE-1$ (LE=undiskontierte Lebenserwartung), wie die folgende Tabelle zeigt.

Lebenserwartung (LE) , 95% Konfidenzintervall, Alter 50 Jahre, Deutschland (2000).				
		Undiskontiert	Aktuariell Diskontiert (5%)	Inkorrekter Ansatz $\sum_{n=0}^{LE-1} 1/(1+0,05)^n$
Männer	schlaganfallfrei	28,13 [28,04; 28,22]	14,80 [14,79; 14,81]	15,68 [15,65; 15,70]
	gesamt	27,45 [27,39; 27,51]	14,64 [14,63; 14,64]	15,49 [15,48; 15,51]
Frauen	schlaganfallfrei	33,25 [33,12; 33,38]	16,19 [16,18; 16,20]	16,85 [16,83; 16,88]
	gesamt	32,43 [32,34; 32,51]	16,03 [16,02; 16,04]	16,68 [16,66; 16,70]

5 Diskussion und Folgerung

Durch Internetsnavigationstechnologie können der Zeitaufwand, die Betriebskosten und die Fehlerrate der Informationsgewinnung aus webbasierten Datenbanken stark reduziert werden. Mittels SAS, MS-PowerPoint, MS-Office Web Komponente und AutoIt kann diskontiertes und undiskontiertes Überleben aus webbasierten Rohdaten in Echtzeit, dynamisch und qualitätsgesichert gewonnen werden.

Literatur

- [1] Ben-Zion B, Reddall R. Life expectancy and actuarial present values: A note to forensic economists. *Research in law and economics* 1985;7:161-171
- [2] Chiang CL. *The life table and its application*. Malabar, Fla, Robert E. Krieger Publishing, 1984m xix, 316 p
- [3] <http://www.gbe-bund.de> (Zugriff am 04.03.2009)

- [4] http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/Statistiken/Bevoelkerung/GeburtenSterbefaelle/Tabellen,templateId=renderPrint.psm1__nnn=true (Zugriff am 04.03.2009)
- [5] SAS onlinedoc with PDF files
<http://www.technion.ac.il/docs/sas9/Welcomepg.pdf> (Zugriff am 04.03.2009)
- [6] Autoit (and Autoitx) v3.3.0 released, 2008 Dec 24
<http://www.Autoitscript.Com/autoit3>
- [7] Office XP Tool: Web Components
<http://www.microsoft.com/downloads/details.aspx?FamilyID=982b0359-0a86-4fb2-a7ee-5f3a499515dd&displaylang=en> (Zugriff am 04.03.2009)
- [8] Stearns D. Programming Microsoft Office 2000 Web Components. Microsoft Press, ISBN:0-7356-0794-X