
Onderzoek naar de correlatie van karakteristieke genoomveranderingen met beroepsgebonden environmentele factoren

Dr. Karel Van Damme (promotor) – 1
Dr. Ludwine Casteleyn – 1
Dr. Lucienne Michaux - 1
Prof. Dr. Anne Hagemeyer - 1
Prof. Dr. Herman Van den Berghe - 2
Dr. Alix Collard - 3

1. Epidemiology and Community Medicine –UIA-
 2. Centrum voor Menselijke Erfelijkheid -KUL-
 3. LAMBDA -FUNDP-
-

INHOUDSTAFEL

I. INLEIDING.....	1
I.1 Algemene context van het onderzoek	
I.2 Doelstellingen van het onderzoek	
II. THEORISCHE CONTEXT.....	2
A. MYELOYDYSPLASTISCH SYNDROOM : ANALYSE VAN HISTORISCHE DATABASE – III/IV/V	
A/III. METHODOLOGIE.....	4
Constructie van de databases	
Resultaten	
Bespreking	
B. MYELOYDYSPLASTISCH SYNDROOM : CASE-REFERENT STUDIE – III/IV/V	
B/III. MATERIAAL EN METHODE	17
Inleidende methodologische beschouwingen	
Bevragingsmethode	
Het registratiedocument of “questionnaire”	
Case definitie	
Control-definitie	
B/IV. RESULTATEN	25
B/V. BESPREKING	26
Tabaksverbruik	
Koffieconsumptie	
Medicamentengebruik	
Blootstelling aan ioniserende straling als gevolg van RX onderzoeken	
Pesticidenblootstelling	
Solventenblootstelling	
Dioxine-blootstelling	
Verband tussen clonale chromosomale veranderingen en pesticidenblootstelling, medicamentengebruik en solventenblootstelling	
CONCLUSIES	31
Evaluatie van de bevragingsmethode	
Blootstellingspecifieke clonale chromosomale veranderingen bij kwaadaardige aandoeningen	
De wenselijkheid van een MDS register	

A. MYELOYDYSPLASTISCH SYNDROOM: ANALYSE VAN HISTORISCHE DATABASE

Annex I	Ruwe database
Annex II	Confirmed databases
Annex III	Extended confirmed database

B: MYELOYDYSPLASTISCH SYNDROOM: CASE-REFERENT STUDIE

Annex IV	Onderzoeksprotocol van de case-referent studie
Annex V	Registratiedocument voor de interviews met cases en controles
Annex VI	Vragenlijsten voor specifieke beroepen: enkele voorbeelden
Annex VII	Analyseverslag

C: ONDERZOEK NAAR SUSCEPTIBILITEITSFACTOREN - VOORSTEL

Annex VIII	Voorstel voor een nieuw onderzoek.
------------	------------------------------------

I. INLEIDING

I.1 Algemene context van het onderzoek

Dit project beoogt onderzoek naar het verband tussen mogelijke environmentele oorzaken van kanker en specifieke genoomaantastingen in de kankercellen, alsook naar de mogelijke relevantie van zulke kennis met betrekking tot de preventie, de herkenning en de (medicolegale) erkenning van beroepsziekten. De resultaten van dit onderzoek zouden in principe moeten kunnen bijdragen tot preventieve acties teneinde de incidentie van kwaadaardige (hematologische) aandoeningen door blootstelling aan scheikundige stoffen - zowel bij werknemers als bij de algehele bevolking - te verminderen. Een beter inzicht in specifieke genoomaantastingen zou een belangrijk element kunnen zijn in de ontwikkeling van testen voor de vroegtijdige opsporing bij werknemers van schadelijke effecten van blootstelling die aanleiding kan geven tot maligne (hematologische) afwijkingen. Het zou ook kunnen helpen om meer accurate methoden te ontwikkelen voor het bepalen van het kankerverwekkend vermogen van bepaalde stoffen, vooraleer deze op de markt worden toegelaten. Bovendien wordt onderzocht of de resultaten van dit onderzoek zouden kunnen bijdragen tot een betere evaluatie van de beroepsomvang van sommige kankers, mede met het oog op eventuele schadeloosstelling door het Fonds voor de Beroepsziekten.

I.2 Doelstellingen van het onderzoek

De specifieke doelstellingen die gezamenlijk werden uitgevoerd door de centra van Leuven, Antwerpen en Namen centreerden zich rond de studie van het Myelodysplastisch syndroom (MDS), en meer in het bijzonder het opzetten en uitvoeren van een retrospectieve analyse van een bestaande gegevensbank en een prospectief epidemiologisch onderzoek.

In nauwe samenwerking tussen de verschillende leden van het netwerk werden in een eerste deel van de studie de karyotypische gegevens bestudeerd van MDS patiënten voor wie de laatste 20 jaar analyses verricht werden in het Centrum voor Menselijke Erfelijkheid .

Dit onderdeel van het onderzoek diende als achtergrond voor het tweede –epidemiologisch-onderzoek van dit project, met name een case-referent studie waarbij de blootstellingsgegevens van nieuwe MDS gevallen die vastgesteld werden in 2 ziekenhuizen gedurende de jaren 1997 en 1998 vergeleken werden met blootstellingsgegevens van referentiepersonen.

De onderzoeksdoelstellingen werden bovendien uitgebreid met de voorbereiding van een onderzoek dat zou moeten toelaten hypothesen te genereren over de susceptibiliteitsfactoren die mogelijk een rol spelen in het ontstaan van environmenteel geïnduceerde hematolymfopoietische aandoeningen enerzijds en van wijzigingen in het perifere bloedbeeld bij matige blootstelling aan benzeen anderzijds. Dit kan tot een beter inzicht leiden in de betekenis van sommige perifere bloedbeeldwijzigingen die worden vastgesteld bij het periodiek

arbeidsgeneeskundig toezicht. Zulke kennis zou tezelfdertijd kunnen bijdragen tot een beter inzicht in de ontstaansmechanismen van hematologische maligniteiten.

Het geheel van deze onderzoeksactiviteiten stemt overeen met de doelstellingen die de centra Antwerpen, Leuven en namen van het onderzoeksnetwerk zich oorspronkelijk hadden gesteld.

II. THEORETISCHE CONTEXT

Het Myelodysplastisch Syndroom (MDS) is een verzamelbegrip voor een groep ernstige hematologische aandoeningen gekenmerkt door een ineffektieve bloedaanmaak met kwantitatieve en kwalitatieve afwijkingen in bloed en beenmerg, die een voorstadium kunnen zijn van acute leukemie: bij ongeveer 30 % van de patiënten met MDS zal de ziekte evolueren naar acute leukemie. In de andere gevallen treedt overlijden na enkele jaren op door onvolwaardige beenmergfunctie.

Betrouwbare gegevens over het voorkomen van MDS zijn niet voorhanden. Men veronderstelt dat de beschikbare incidentie cijfers een onderschatting zijn van de realiteit. In de voorbije decennia werd de diagnose van MDS steeds vaker gesteld, mede dankzij een betere diagnostiek, en clinici hebben de -al dan niet terecht- indruk dat de frequentie van deze aandoening die zich vooral voordoet bij ouderen, ook is toegenomen in jongere leeftijdscategorieën.

Epidemiologische onderzoeken naar het verband met uitwendige blootstelling zijn zeldzaam. Enkele studies toonden of suggereerden reeds een verband tussen het Myelodysplastisch Syndroom en professionele of environmentele blootstelling aan organische stoffen, aan anorganische gassen, aan ioniserende en niet-ioniserende straling en aan metalen. Epidemiologisch onderzoek kan zulke correlaties aantonen, doch dit houdt niet noodzakelijk in dat er een rechtstreeks oorzakelijk verband is.

Het bijzondere aan MDS is dat het ons wellicht méér dan andere aandoeningen toelaat om oorzakelijke verbanden met uitwendige factoren aan te tonen. Dit syndroom komt namelijk frequent voor als gevolg van een behandeling met chemotherapeutica, en in vele van deze gevallen dragen de aangetaste cellen bepaalde kenmerkende chromosomale veranderingen die dan ook als de handtekening beschouwd kunnen worden van een uitwendige oorzaak. De meest bekende van deze chromosomale wijzigingen is 5q- en werd ontdekt in het Centrum voor Menselijk Erfelijkheid te Leuven. Identieke veranderingen vindt men ook terug bij MDS patiënten die géén behandeling met chemotherapeutica hadden ondergaan. Alerten clinici hadden de stellige indruk dat deze patiënten intens waren blootgesteld aan bepaalde uitwendige factoren. Indien uit onderzoek zou blijken dat er een verband bestaat tussen deze chromosomale afwijkingen en uitwendige oorzaken, en indien deze oorzaken geïdentificeerd kunnen worden, zou dit een bijzondere stap kunnen betekenen niet enkel voor de preventie van MDS, doch ook voor het ontwikkelen van methodes van gezondheidsbewaking. Omwille van al deze redenen is de studie van het Myelodysplastisch syndroom bijzonder aangewezen in het kader van een

onderzoek naar blootstellings specifieke aantastingen van het genoom. Daarnaast zou deze kennis inzake de associatie tussen bepaalde uitwendige factoren en bepaalde aantastingen van het genoom interessante inzichten kunnen opleveren voor het fundamentele biogenetisch onderzoek.

Behalve 5q- zijn er nog andere clonale chromosomale veranderingen die bij MDS kunnen worden aangetroffen en waarvan wordt aangenomen dat ze verband houden met uitwendige blootstelling. Een eerste betrachting in dit onderzoek was om een volledig beeld te bekomen van de verschillende clonale chromosomale veranderingen bij MDS en de frequentie waarmee ze worden aangetroffen. Dit werd bestudeerd door de onderzoeksteams aan de hand van een historische en bijzonder uitgebreide database bij het Centrum voor Menselijke Erfelijkheid aan de KUL. Dit onderzoek vormt het eerste luik van dit rapport.

Een tweede luik, wordt gevormd door een case-referent onderzoek, en de daaropvolgende analyse van het verband tussen significante blootstellingsfactoren en de clonale chromosomale afwijkingen bij cases. Het case-referent onderzoek is een noodzakelijke tussenstap in de verdere studie van de blootstellings specifieke aantastingen van het genoom. Hiervoor dient het beperkte onderzoeksmateriaal met een zo gevoelig mogelijke methode te worden benaderd, vooreerst teneinde mogelijke causale factoren zo precies mogelijk op te sporen. Doch er is een tweede en zeer belangrijke reden waarom deze nauwkeurigheid in de reconstructie van de blootstellingsgeschiedenissen een absoluut streven dient te zijn dat voorrang dient te hebben op andere consideraties zoals 'standaardisering' van de bevraging: het streven naar de uiterste nauwkeurigheid is noodzakelijk indien men voor de cases het verband wenst te exploreren tussen deze blootstellingsfactoren en de clonale chromosomale veranderingen in de aangetaste cellen. Daar waar gedeeltelijke foute informatie niet noodzakelijk tot foute conclusies zal leiden in de vergelijking tussen cases en referents indien de foute informatie gelijk gespreid is over de antwoorden van cases en referents, zal dezelfde foute informatie met zekerheid tot foute conclusies leiden wanneer ze gebruikt wordt om de blootstellingsgeschiedenissen van cases met bepaalde chromosomale afwijkingen te vergelijken met deze van andere cases zonder deze afwijkingen. De onderzoeksteams hebben dan ook getracht om een epidemiologische methode te ontwikkelen die zeer precieze en dus genuanceerde informatie oplevert over de blootstellingsgeschiedenissen van cases en referentiepersonen. Deze methode werd toegepast voor de studie van MDS. In afzonderlijke bijlagen bevinden zich het onderzoeksprotocol (annex IV), de bijhorende registratiedocumenten of vragenlijsten (annexen V en VI) en de onderzoeksresultaten of analyseverslag (annex VII). In dit rapport worden al deze elementen toegelicht.

Het onderzoeksproject werd bovendien uitgebreid met de voorbereiding van en bijkomend onderzoek naar specifieke susceptibiliteitsfactoren die mogelijk een rol spelen in het ontstaan van environmenteel geïnduceerde hematolymfopoietische aandoeningen. Van deze genen is bekend of wordt aangenomen dat zij tussenkomen bij het metabolisme van lichaamsvreemde stoffen. Hun polymorfisme verklaart mogelijks een deel van de grote interindividuele variabiliteit in gezondheidseffecten bij blootstelling aan toxische stoffen.

Voor verdere achtergrondbeschouwingen verwijzen we naar de inleidende hoofdstukken van het onderzoeksprotocol van de case-referent studie.

A. MYELODYSPLASTISCH SYNDROOM: ANALYSE VAN HISTORISCHE DATABASE – III/IV/V

A/III METHODOLOGIE

Constructie van de databases

De CME database

In het Centrum voor Menselijke Erfelijkheid te Leuven worden de karyotypes onderzocht van patiënten van hematologische en oncologische diensten van hospitalen uit België en het Groot-Hertogdom. Bij de sedert 1975 opgeslane gegevens wordt vermeld welke de door de clinicus opgegeven indicatie voor het onderzoek was (rubriek 'comment'). In sommige gevallen werd bij de mededeling van het onderzoeksresultaat aan de clinicus een formulier bijgevoegd waarbij meer inlichtingen werden gevraagd over de aandoening van de patiënt. Indien dit formulier 'vraag om inlichtingen' werd teruggestuurd met de gevraagde gegevens werd de opgegeven diagnose als een apart gegeven bij de 'comments' gevoegd.

De database werd destijds niet opgezet met de bedoeling om direct bruikbaar materiaal ter beschikking te stellen voor epidemiologisch onderzoek. De vraag stelde zich derhalve hoe deze uitzonderlijke verzameling gegevens voor onze onderzoeksdoelstellingen kon worden gebruikt. De gevolgde stappen worden hieronder weergegeven.

Van CME database naar 'ruwe MDS database'

Uit de bestaande database van gegevens verzameld tussen 1/1/75 en 1/1/97 werden die patiënten geselecteerd waarbij de door de clinicus opgegeven reden voor onderzoek indicatief werd geacht voor een diagnose van MDS. Het bepalen van de lijst van indicaties die zou aanleiding geven tot inclusie in de database gebeurde in verschillende stappen. Navraag bij verschillende experts uit de kliniek en uit het labo gaf een eerste lijst van inclusiecriteria. De lijst van patiënten uit de CME database die beantwoordden aan die criteria vormde een eerste 'ruwe MDS database'. Deze werd vervolgens vergeleken met een lijst van patiënten waarbij de diagnose van MDS was bevestigd via het formulier 'vraag om inlichtingen' en dus los van de opgave van indicatie voor onderzoek (zie boven). Een aantal van deze patiënten kwamen niet voor in de eerste 'ruwe MDS database' en de algemene inclusie criteria werden aangepast om dit te verhelpen.

Vervolgens werden de data nagezien op administratieve onnauwkeurigheden en verbeterd waar nodig.

Nadien werd gepoogd de database verder te verfijnen: voor een aantal gevallen werd de diagnose MDS als onwaarschijnlijk geacht of aanzien als niet te weerhouden, omdat (1) er een duidelijke andere diagnose aan te wijzen was via gegevens naderhand verkregen bij de clinicus (2) er slechts 1 observatiedatum was betreffende die patiënt, de daarbij opgegeven indicatie zeer specifiek en het bekomen karyotype volledig normaal. Deze (vermoedelijk) 'vals +' gevallen werden niet behouden in de database.

Uiteindelijk werd aldus een 'ruwe MDS database' weerhouden met gegevens over 8274 observaties betreffende een totaal van 4134 patienten.

Van 'ruwe MDS database' naar 'confirmed MDS database'

Aangezien mag verwacht worden dat in deze 'ruwe MDS database' nog patienten ingesloten zijn die niet aan MDS leden, werd vervolgens een selectie gemaakt uit de ruwe MDS database waarin alleen die patienten voorkomen waarbij de diagnose MDS zo goed als zeker is:

- (1) de patienten waarvoor de behandelend ziekenhuisarts een 'vraag om inlichtingen' formulier invulde met als diagnose myelodysplastisch syndroom
- (2) de patienten die voorkwamen op de lijsten van bevestigde MDS diagnoses. Deze lijsten werden verkregen van de artsen van 3 van de verwijzende centra.

Op deze manier werd een 'confirmed database' bekomen van 602 patienten met samen 1727 observaties.

Van 'confirmed MDS database' naar 'extended confirmed database'

De lijsten met MDS patienten verkregen van de 3 verwijzende centra lieten ons ook toe de sensitiviteit van onze selecties (databases) te beoordelen. Een aanzienlijk aantal patienten vermeld op de lijst van de centra kwamen niet voor in onze 'ruwe database' en dus ook niet in onze 'confirmed database'. Via nazicht van individuele gegevens in de CME database werd gepoogd de redenen hiervoor op te sporen:

- (1) een aantal van deze patienten zijn niet opgenomen in het CME bestand;
- (2) een aantal van hen komen wel voor in het CME bestand, maar er werd bij geen enkel staal een indicatie voor onderzoek, een 'comment', aangegeven, zodat deze patienten met de gehanteerde inclusiecriteria vanzelfsprekend niet voorkomen in de ruwe MDS database;
- (3) bij een derde groep werd wel een comment aangegeven, maar deze is zo weinig specifiek voor het Myelodysplastisch Syndroom dat deze niet als inclusiecritrium kan gebruikt worden.

In een laatste stap in de behandeling van de gegevens werd dan de 'confirmed database' aangevuld met de gegevens van de patienten die vermeld zijn onder (2) en (3) hierboven.

Omdat de diagnostische criteria voor 1980 minder duidelijk omschreven waren werden alle gevallen verwijderd waarbij de diagnose voor 1980 werd gesteld. Begin jaren 80 werd de French American British classification (Bennet et al, 1982) geleidelijk aan algemeen gehanteerd.

Ook alle gevallen van patiënten jonger dan 15 jaar werden uitgesloten. Aldus werd een bestand bekomen van 2412 observaties over 793 patiënten.

Database voor analyse van karyotypes

Voor de analyse van de karyotypes werd uit de 'extended confirmed database' een selectie gemaakt met (1) uitsluiting van de 97 patiënten waarbij geen gegevens over het karyotype werden bekomen; (2) uitsluiting van de 43 patiënten waarbij nooit resultaten van beenmerg onderzoek werden bekomen en waarbij eveneens nooit karyotypische afwijkingen op de perifere bloedcellen werden vastgesteld.

Voor analyse van de cytogenetische gegevens van MDS is beenmerg immers het aangewezen materiaal. Dergelijk materiaal is echter niet altijd beschikbaar. Bovendien kunnen resultaten van een karyotypering van het beenmerg ook ontbreken om technische redenen. Wanneer in dergelijke gevallen een perifere bloedcel onderzoek een abnormaal karyotype opgeeft hebben we een aanwijzing voor de afwijkingen in het beenmerg. Omgekeerd echter is een normaal karyotype in de perifere bloedcellen geen aanwijzing voor een normaal karyotype in het beenmerg.

Een bestand 653 patiënten was aldus beschikbaar voor verdere analyse.

A/IV. RESULTATEN

Beschrijving van de verzamelde gegevens

'Ruwe database' (Zie Annex I voor volledig rapport)

De 'ruwe MDS database' bevat gegevens over 8274 observaties betreffende een totaal van 4134 patiënten. De relevante informatie beschreven dmv 16 variabelen wordt samengevat in onderstaande tabel

Variable list for the MDS retrospective study database.

Variable name	Variable content
Patid	Patient identification number
Sex	Sex
Zip	zip code (residence)
Age	Age
Birth.date	birth date
Sample.num	Sample number
Sample.date	Sample date
Inst	Hospitale
Tissue.1	tissue type
Tissue.2	Sample number for tissue type specified in Tissue.1
Diag	Diagnostic
Lab	lab number
Clone	clone number
Number.cell	Number of cells within one clone
Karyo	Karyotype
Status	status code

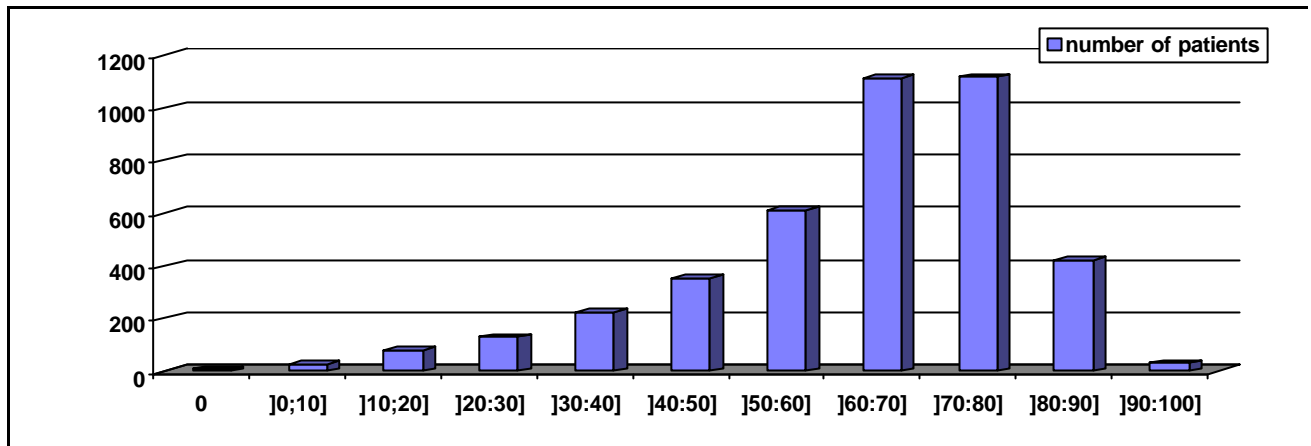
Een 17de variabele vermeldt de gegevens van het confirmatie-formulier zoals hoger beschreven.

Deze populatie bestaat uit 2 172 (52.5%) mannen en 1 946 (47.1%) vrouwen. Voor 16 personen waren geen gegevens over het geslacht beschikbaar. Deze sex ratio van 1.1 is significant verschillend van 1 (p-value = 0.0002). In de literatuur vindt men over het algemeen een (nog duidelijker) overwicht van mannen onder MDS patienten. 785 patienten (19 %) van de globale groep zijn jonger dan 50 jaar. 1847 (44.7%) patienten zijn jonger dan 65 jaar. De gemiddelde leeftijd is 62.89 jaar. Deze selectie van patienten is jonger dan mag verwacht worden uit de literatuur gegevens voor MDS patienten.

Summary of age distribution for population database.

Number of patients	4134
Number of NA	22
Mean	62.89
Standard deviation	16.89
Minimum	0
Median	66
Maximum	99

Age distribution for patients in the global database.



580 Mannen en 392 vrouwen uit deze populatie vertonen een duidelijk abnormaal karyotype. Dit is samen 19,4 % van de gehele populatie. Hun gemiddelde leeftijd is zowat 64,3 jaar. De sex ratio is 1.5.

De 'confirmed database' (Zie Annex II voor volledig rapport)

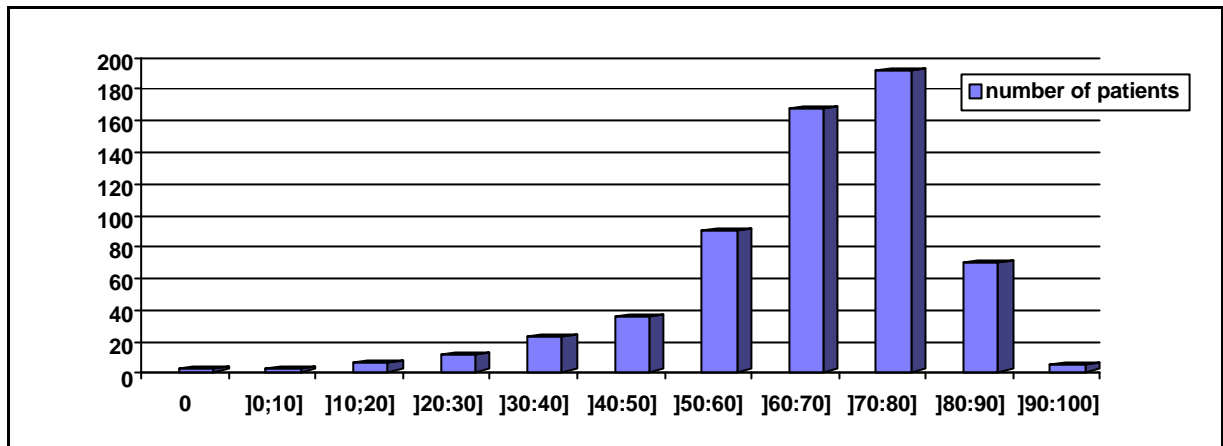
De 'confirmed database' betreft 602 patienten met samen 1727 observaties. Deze populatie bestaat uit 351 mannen (58.3%) en 248 vrouwen (41.7%). Voor 3 patienten was geen informatie bekend over het geslacht. De sex ratio is hier dus 1.41, significant verschillend van 1 (p-value < 0.0001).

74 patienten (12.3%) zijn jonger dan 50 jaar en 222 (36.9%) jonger dan 65 jaar. De gemiddelde leeftijd is 65.83 jaar.

Summary of age distribution for population database.

Number of patients	602
Number of NA	4
Mean	65.83
Standard deviation	15.09
Minimum	0
Median	69
Maximum	99

Age distribution for patients in the global database.



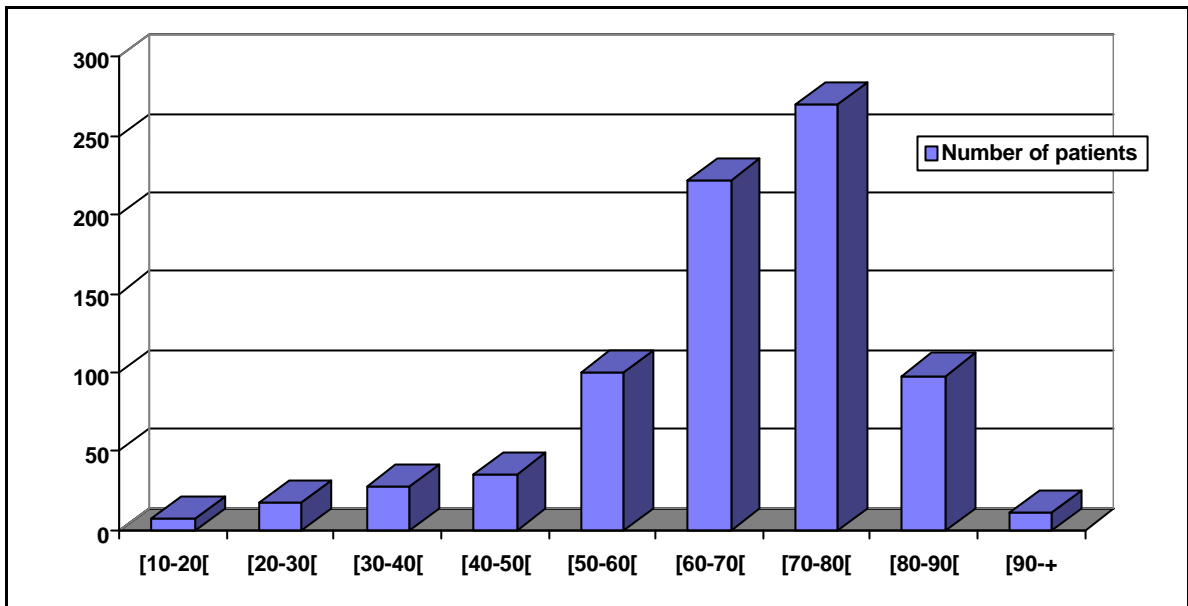
De 'extended confirmed MDS database' (Zie Annex III voor volledig rapport)

De 'extended confirmed database' betreft 793 patiënten met samen 2412 observaties. Deze populatie bestaat uit 466 mannen (58.8%) en 324 vrouwen (40.8 %). Voor 3 patiënten was geen informatie bekend over het geslacht. De sex ratio is hier dus 1.42, significant verschillend van 1 (p-value < 0.0001).

Op het moment van de eerste observatie zijn 97 (12.2%) van deze patiënten jonger dan 50 jaar en 304 (38.1%) jonger dan 65 jaar. De gemiddelde leeftijd is 66.47 jaar.

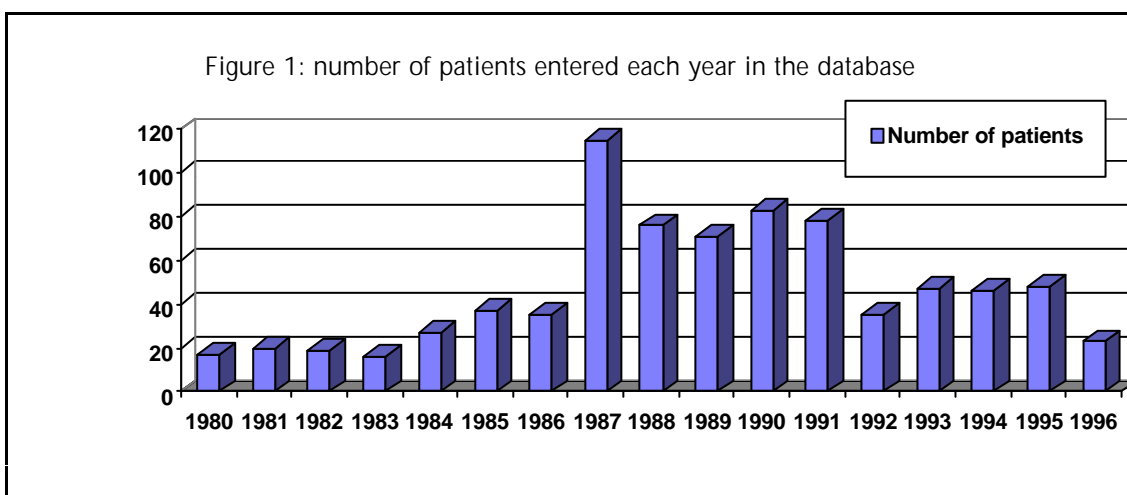
Summary of age distribution

Number of patients	793
Number of NA	5
Mean	66.47
Standard deviation	14.36
Minimum	15
Median	69
Maximum	100



Age distribution

Het aantal patiënten met een eerste observatie is per jaar weergegeven in de volgende figuur. De aanzienlijk hogere cijfers gedurende de jaren 1987 tot 1991 worden verklaard door de specifieke strategie die gevolgd werd bij het samenstellen van deze database. Zoals vermeld onder punt 1 (constructie van de database) werden een groot aantal van de gevallen immers in deze database opgenomen omwille van het feit dat de diagnose door de behandelend ziekenhuisarts geconfirmeerd werd aan het labo via een specifiek formulier. In de bewuste piekperiodes werden deze formulieren verspreid en op een vrij systematische manier gebruikt. Nadien raakten deze documenten terug wat in onbruik.



Database voor analyse van karyotypes (Zie Annex III, tweede deel)

Van de 653 patiënten in deze database vertonen er 272 minstens 1 observatie met een abnormaal karyotype. Bij 381 patiënten werd nooit een abnormaal karyotype vastgesteld. De constitutionele afwijking t(13;14), die bij 3 patiënten gevonden werd, werd buiten beschouwing gelaten. De populatie bestaat uit 264 vrouwen en 389 mannen.

De volgende tabellen tonen aan dat de abnormale karyotypes +/- gelijk verdeeld zijn over mannen en vrouwen. Patiënten met een karyotypische afwijking zijn significant jonger dan deze met een normaal karyotype ($p=0.0334$). Wanneer men de leeftijdsdistributie per geslacht beschouwt blijkt dat dit verschil enkel bij de vrouwen blijft bestaan ($p = 0.0369$) en bij de mannen uitgevlakt wordt ($p = 0.3179$).

Comparison of sex repartition across patients with and without anomaly:

Sex	With anomaly	Without anomaly	P-value*
Female	107 (39.3%)	157 (41.2%)	
Male	165 (60.7%)	224 (58.8%)	0.6899
	272 (100%)	381 (100%)	

* chi-square test

Age comparison across patients with and without anomaly:

Age characteristics	With anomaly	Without anomaly	P-value*
Number of patients(number of NA)	272 (1)	381 (2)	
Mean \pm sd	64.27 \pm 16.07	67.30 \pm 13.28	
Min-med-max	15-67-100	15-70-93	0.0334

* Wilcoxon rank-sum test

Age characteristics across patients with and without anomaly, by sex

Age characteristics	With anomaly	Without anomaly
		Female
N (na)	107 (0)	157 (1)
Mean \pm sd	63.21 \pm 16.07	67.97 \pm 14.40
Min-med-max	16 -68-97	15-71-91
		Male
N (na)	165 (1)	224 (1)
Mean \pm sd	64.97 \pm 14.98	66.84 \pm 12.44
Min-med-max	15-67-100	19-69-93

Male versus female comparison: within patients with anomaly: 0.8076
within patients without anomaly: 0.1271

Patients with anomaly versus patients without anomaly: within female patients: 0.0396
within male patients: 0.3179

(All tests for the above table are Wilcoxon rank-sum tests)

De populatie in deze database kan verder ingedeeld worden in een groep met normale karyotypes, een groep met een afwijkend karyotype waarbij maximaal ter hoogte van 2 chromosomen afwijkingen gevonden worden (simple), en tenslotte een groep met een complex

abnormaal karyotype waarbij ter hoogte van 3 of meer chromosomen een afwijking gevonden wordt (complex).

Age characteristics	Complex karyotype	Simple karyotype	Normal karyotype
Sex F/M	33 / 40	74 / 125	157 / 224
Number of patients(number of NA)	73 (1)	199 (0)	381 (2)
Mean \pm sd	63.79 \pm 15.56	64.45 \pm 16.28	67.30 \pm 13.28
Min-med-max	16 - 66.5 – 85	15 - 67 – 100	15 - 70 – 93
p-value*	0.0878	0.0847	

* Wilcoxon rank-sum test for the comparison towards the patients with a normal karyotype

Onder de 73 patienten met een 'complex' karyotype vinden we 21 (28.76%) patienten waarbij geen enkele afwijking gevonden wordt ter hoogte van chromosoom 5 en 7. Deze karyotypes zonder afwijkingen ter hoogte van 5 en 7 vinden we vooral terug bij de groep waar niet meer dan 3 chromosomen aangetast zijn. Wanneer we de groep beschouwen met afwijkingen ter hoogte van minstens 4 chromosomen dan dragen 49 van de 59 cases ook afwijkingen ter hoogte van 5 en 7.

Voor het bepalen van een relevante subgroep met 'complex karyotype', refererend naar de karyotypes die vaak aangetroffen worden bij secundaire MDS en die een slechte prognose inhouden, wordt bijgevolg best als criterium 'ten minste afwijkingen ter hoogte van 4 chromosomen' gehanteerd.

De verdere analyse betreft enkele specifieke chromosoomafwijkingen die als karakteristiek beschouwd voor secundaire MDS of AML (acute myeloïde leukemie) te wijten aan vroegere chemotherapie met alkylerende agentia of die men ook vaak vindt bij patienten met een geschiedenis van beroepsmatige blootstelling aan organische solventen of insecticiden.

CHROMOSOOM 5

73 patienten vertonen minstens 1 chromosomale afwijking ter hoogte van chromosoom 5. Onder hen vinden we 43 vrouwen en 30 mannen. De gemiddelde leeftijd is 65.19 ± 15.24 jaar.

Bij analyse van de eerste observatie met een afwijking ter hoogte van chromosoom 5 vinden we 26 patienten met « 5q- » als enige afwijking, 21 patienten waarbij deze 5q- voorkomt in combinatie met andere afwijkingen, en 18 patienten met « -5 ». De overige 8 patienten vertonen diverse andere afwijkingen ter hoogte van chromosoom 5. Geslachts- en leeftijdsdistributie worden gegeven in de volgende tabel.

	5q- alone	5q- plus other	-5	other 5
Sex: F / M	19 / 7	11 / 10	9 / 9	4 / 4
Age Nb (Nb of NA)	26 (1)	21 (0)	18 (0)	8 (0)
Mean \pm sd	64.60 \pm 17.45	66.05 \pm 11.71	66.72 \pm 14.01	61.37 \pm 20.46
Range	20 - 67 - 85	40 - 65 - 83	24 - 65 - 84	18 - 71 - 79

CHROMOSOOM 7

69 patiënten vertonen minstens 1 chromosomale afwijking ter hoogte van chromosoom 7. Onder hen vinden we 32 vrouwen en 37 mannen. De gemiddelde leeftijd is 60.13 ± 17.95 jaar.

Bij analyse van de eerste observatie met een afwijking ter hoogte van chromosoom 7 vinden we 11 patiënten met « 7q- » en 48 patiënten met « -7 ». De overige 10 patiënten vertonen diverse andere afwijkingen ter hoogte van chromosoom 7. Geslachts- en leeftijdsdistributie worden gegeven in de volgende tabel.

	7q-	-7	Other
Sex: F / M	1 / 10	24 / 24	7 / 3
Age Nb (Nb of NA)	11 (0)	48 (0)	10 (0)
Mean \pm sd	66.36 \pm 13.37	56.79 \pm 19.28	69.30 \pm 9.65
Min-med-max	35 - 65 - 85	16- 60 - 84	52 - 68.5 - 83

-7 is bekend als een afwijking die voorkomt op jongere leeftijd. In de bestudeerde populatie zijn patiënten die bij de eerste observatie een -7 vertonen, inderdaad aanzienlijk jonger (gemiddelde leeftijd 56.79 jaar). In de literatuur worden 7q- en -7 vaak samen bekeken, waardoor deze typische leeftijdsdistributie uitgevlakt wordt.

CHROMOSOOM 8

67 patiënten vertonen minstens 1 chromosomale afwijking ter hoogte van chromosoom 8. Onder hen vinden we 22 vrouwen en 45 mannen. De gemiddelde leeftijd is 65.78 ± 14.99 jaar.

Bij analyse van de eerste observatie met een afwijking ter hoogte van chromosoom 8 vinden we 57 patiënten met « +8 » en 10 patiënten met diverse andere afwijkingen ter hoogte van chromosoom 8. Geslachts- en leeftijdsdistributie worden gegeven in de volgende tabel.

	+8	Other
Sex: F / M	19 / 38	3 / 7
Age Nb (Nb of NA)	57 (0)	10 (0)
Mean \pm sd	66.5 \pm 15.63	64.20 \pm 11.21
Min-med-max	18 - 69 - 87	44 - 67 - 80

CHROMOSOOM 11

Negentien patiënten hebben een afwijking ter hoogte van chromosoom 11 zonder dat daarbij ook chromosoom 5 en/of 7 aangetast zijn. Het gaat om 7 vrouwen en 12 mannen met de volgende leeftijdsdistributiegegevens: [19 (0) ; 67.84 ± 13.34 ; 33 - 68 - 87]

De afwijkingen ter hoogte van chromosoom 11 houden vooral een verlies in van een deel van chromosoom 11. We vinden inderdaad 9 maal een deletie van 11q en 6 maal een verlies van het volledige chromosoom 11. Deze delieite komen op een na zonder afwijking ter hoogte van chromosoom 5 en 7 voor. Verlies van chromosoom 11 of een deel ervan lijkt dus een onafhankelijke rol in de pathogeneses van MDS te spelen.

CHROMOSOOM 12

De afwijkingen ter hoogte van chromosoom 12 bevinden zich bij 24 gevallen ter hoogte van de korte arm en bij 8 gevallen ter hoogte van de lange arm. In 2 gevallen zijn zowel de korte als de lange arm aangetast. Vooral de korte arm lijkt dus van belang te zijn in de pathogenese van MDS.

CHROMOSOOM 20

44 patiënten vertonen minstens 1 chromosomale afwijking ter hoogte van chromosoom 20. Onder hen vinden we 15 vrouwen en 29 mannen. De gemiddelde leeftijd is 64.74 ± 13.72 jaar.

Bij analyse van de eerste observatie van de betreffende afwijking vinden we de 29 patiënten met « 20q- ». De overige 10 patiënten vertonen diverse andere afwijkingen ter hoogte van chromosoom 20. Geslachts- en leeftijdsdistributie worden gegeven in de volgende tabel.

	20q-	-20	Other
Sex: F / M	10 / 19	2 / 3	3 / 7
Age Nb (Nb of NA)	29 (0)	5 (0)	10 (1)
Mean ± sd	63.41 ± 13.85	66.20 ± 9.50	68.22 ± 15.82
Range	31 – 83	54 – 76	43 – 83

A/V. BESPREKING

1. Daar de oorspronkelijke ruwe database duidelijk onvoldoende specificiteit vertoonde was het niet mogelijk van deze ruwe gegevens te vertrekken voor de studie van MDS.

Uit de exploratie van deze database kon wel geleerd worden welke voorwaarden vervuld moeten zijn om tot een accurate direct bruikbare MDS registratie te komen die toelaat inzicht te verwerven in het voorkomen van deze aandoening en in de genetische afwijkingen die ermee

gepaard gaan. Zulke registratie moet in de eerste plaats gebaseerd zijn op diagnoses die gesteld zijn volgens duidelijk omschreven criteria en die onderworpen zijn aan een systematische kwaliteitscontrole. Zij moet tevens genetische gegevens bevatten over de aangetaste cellen. Deze gegevens zijn belangrijk voor onderzoek in verband met behandeling, prognose en etiologie. Koppeling aan een centraal georganiseerde bewaring van beenmerg en bloedstalen zou ideaal zijn en moet toelaten om telkens als het nodig is terug te grijpen naar oorspronkelijk biologisch materiaal in het licht van nieuwe inzichten of nieuwe diagnostische middelen.

Daar patiënten met MDS zowel in hematologische als in oncologische diensten gevolgd worden moeten beide betrokken worden in de opbouw van een hospital-based registratie.

Voor verder onderzoek naar externe factoren in de etiologie van MDS is het van essentieel belang dat de gegevens in een MDS register op directe en eenvoudige wijze identificeerbaar zijn, en dus onmiddellijk identificerend kunnen gemaakt worden voor geautoriseerde onderzoekers. Het systeem van de huidige kankerregisters in ons land waarbij enkel de notifiërende arts de terugkoppeling kan doen laat niet toe dat bvb een individuele bevraging van alle patiënten behorend tot één cluster gebeurt wanneer dit wenselijk zou zijn om het de invloed van milieufactoren te bestuderen.

2. Wat betreft de 'extended confirmed database' mag daarentegen aangenomen worden dat de opgenomen observaties inderdaad MDS patiënten betreffen.

Over (trends in) de incidentie van de aandoening in een populatie kan echter niets afgeleid worden daar de catchment populatie wisselend is en onbekend. Het aantal geregistreerde gevallen staat dan ook niet in een systematische verhouding staat tot de werkelijk gediagnosticeerde gevallen.

Men mag wel aannemen dat de registratie niet beïnvloed werd door de leeftijd van de patient. De leeftijdsdistributie van deze populatie (gemiddelde leeftijd 66.46 ± 14.36) (diagnoses voornamelijk vanaf 87) volgt de trend die men in de literatuur van de laatste jaren terugvindt. De studie van C. Nisse (Nisse, 1996) vermeldt bijvoorbeeld een gemiddelde leeftijd van 67.3 ± 12.5 jaar (p-value for two-sided comparison of both sample age characteristics: 0.47). Andere groepen vonden een gemiddelde leeftijd van ongeveer 70 jaar (Third MIC Co-operative study group, 1987) en van 73 jaar (Mufti et al, 1985). Algemeen wordt aangenomen dat meer dan 75% van de patienten ouder zijn dan 60 jaar bij diagnose (Aul et al, 1995). In onze populatie zijn slechts 58.3 (73.1%) patienten ouder dan 60 op het ogenblik van de eerste observatie.

Gemiddelde leeftijd

*	Nisse, 96	MIC, 1987	Mufti et al, 1985
66	67	70	73

3. De analyse van de karyotypes toont in de eerste plaats aan dat de abnormale karyotypes +/- gelijk verdeeld zijn over mannen en vrouwen, maar dat patienten met een karyotypische afwijking significant jonger zijn dan deze met een normaal karyotype ($p=0.0334$). Wanneer men de leeftijdsdistributie per geslacht beschouwt blijkt dat dit verschil enkel bij de vrouwen blijft bestaan ($p = 0.0369$) en bij de mannen uitgevlakt wordt ($p = 0.3179$).

Voor het bepalen van een relevante subgroep met 'complex karyotype', refererend naar de karyotypes die vaak aangetroffen worden bij secundaire MDS en die een slechte prognose inhouden, wordt als criterium 'ten minste afwijkingen ter hoogte van 4 chromosomen' voorgesteld.

Verder valt op dat voornamelijk chromosomen 11 en 12p onafhankelijk van chromosomen 5 en 7 een rol lijken te spelen in de pathogenese van MDS. Deze chromosomen zouden dus een target moeten zijn voor meer gespecialiseerde onderzoeken (Fish, ...) die toelaten ook afwijkingen aan te tonen die niet met klassieke cytogenetische technieken kunnen worden gedetecteerd.

B. MYELODYSPLASTISCH SYNDROOM: CASE-REFERENT STUDIE – III/IV/V

B/III. MATERIAAL EN METHODE

(zie tevens **annex IV**: onderzoeksprotocol)

Inleidende methodologische beschouwingen

Het hier gerapporteerde onderzoek is een case-referent of case-controle onderzoek. Samen met case-series en cohorte-onderzoek vormt deze onderzoeksmethode de belangrijkste bron van onze kennis inzake kankerverwekkende stoffen bij de mens. Bij cohorte-onderzoek naar oorzakelijk factoren van ziekten kan men zich in principe beperken tot het nagaan van het verband tussen één blootstellingsfactor en de bestudeerde aandoening. Bij cohorte-onderzoek dient men erover te waken dat men voldoende grote groepen blootgestelden onderzoekt en een zo adequaat mogelijk geselecteerde vergelijkingsgroep, die in theorie enkel verschilt van de groep blootgestelden door blootstelling aan de bestudeerde factor. Bij case-controle onderzoek dient men –welke ook de factor is die de bijzondere interesse van de onderzoekers wegdraagt– steeds een hele reeks blootstellingsfactoren na te gaan. Vooraleer uit een case-controle onderzoek te besluiten dat een gevonden correlatie tussen bvb een bepaald beroep en een bepaalde aandoening een aanduiding van oorzakelijkheid zou kunnen geven, dient men andere factoren uit te sluiten en derhalve de blootstelling te bestuderen aan andere factoren die tot de categorie life-style of leefomgeving of medisch verleden behoren, en die tot dezelfde aandoening aanleiding zouden kunnen geven. Zoniet zou een dergelijk onderzoek enkel tot volstrekt oninterpreteerbare of onbetrouwbare conclusies leiden. Het is derhalve noodzakelijk om de cases en controlepersonen een hele reeks vragen te stellen. De evenzovele antwoorden dienen een juiste weergave te geven van de feiten, hetgeen een groter probleem kan vormen naarmate het interview langer duurt, des te meer daar het hier veelal gaat over ernstig zieken.

Het verzamelen van correcte en uitvoerige informatie over het blootstellingsverleden is dus de sleutelopdracht bij case-referent onderzoek naar mogelijk oorzakelijke verbanden met ziekten. Meestal gebeurt dit door bevraging van slachtoffers van de bestudeerde ziekte enerzijds en referentiepersonen die niet aan de ziekte lijden anderzijds, gevolgd door het vergelijken van het blootstellingsverleden van beide groepen personen. Deze bevraging kan gevolgd worden door het verzamelen van verdere informatie via verschillende kanalen.

Het doel van de bevraging van patiënten (cases en controls of referents) moet zijn om zo precies mogelijke antwoorden te bekomen op relevante vragen. Daarnaast moet de bevraging zo gebeuren dat zij een betrouwbare vergelijking toelaat van de antwoorden die werden gegeven respectievelijk door de cases en de controles of referentiepersonen. Het is immers uit deze vergelijking dat aanduidingen voor mogelijke oorzakelijke verbanden worden gehaald.

Aan deze drievoudige vereiste (relevante vragen, juiste antwoorden en vergelijkbaarheid) kan nooit op een perfecte manier tegemoet gekomen worden. Zij is het voorwerp van steeds weerkerende methodologische discussies.

Ons inziens is het een ietwat paradoxale doch ruim verbreide gewoonte in epidemiologische kringen om in deze discussies de hoofdaandacht te besteden aan de vergelijkbaarheid van de antwoorden, en daarbij niet altijd de nodige aandacht te hebben voor het feit dat de juistheid van de antwoorden wel eens zou kunnen lijden onder de strategie voor de vergelijkbaarheid. Daarbij is het bevragingparadigma een haast gerobotiseerde situatie die als volgt kan omschreven worden: een daartoe opgeleid persoon ondervraagt cases en controles zonder te weten wie de cases en de controles zijn. De goed bestudeerde vragen worden op een gestandaardiseerde wijze gesteld, en de antwoorden op een gestandaardiseerde wijze genoteerd. Daarbij mag de ondervrager zich geen enkele interpretatie veroorloven (om elke interviewer bias te vermijden, indien deze toch weet zou hebben van de case status of controle status van de geïnterviewde). Men gaat er verder van uit dat –indien er geen sprake is van systematische recall bias (waarbij verondersteld wordt dat de cases eerder zullen geneigd zijn om blootstellingen aan te halen of ze te overdrijven dan de controles) - de kans dat de vragen onjuist worden beantwoord gelijk is voor cases en controles (hetgeen des te meer het geval zal zijn naarmate de onderzochte aantallen groter zijn) en dat de eventuele onjuiste antwoorden van geen invloed zouden zijn op de onderzoeksresultaten daar zij elkaar neutraliseren. Dit laatste lijkt ons een veeleer optimistische dan onderbouwde gedachte. Grotere bezwaren lijken ons dat een streng vormelijk gestandaardiseerde benadering niet toelaat om recall bias te herkennen en te corrigeren, en dat relevante informatie verloren kan gaan wanneer zowel cases als controles op een systematische manier dezelfde verkeerde inschattingen maken van hun blootstellingsverleden.

Het verzamelen van relevante en dus juiste informatie zowel bij cases als controles is de hoofdvereiste. Vanuit deze bekommernis dient de epidemiologische methode gestuurd te worden, en niet vanuit een allesoverheersende vrees voor subjectiviteit bij de bevraging. Een discours dat de subjectiviteit van de onderzoekers als het grootste gevaar beschouwt van een epidemiologisch onderzoek, blokkeert een aantal pistes voor het oplossen van het ons inziens onderschatte probleem van de foutieve informatie.

Foutieve informatie zou ons onderzoek voor een bijkomende en onoverkomelijke handicap stellen. Zoals hoger uiteengezet wordt de identificatie van mogelijks oorzakelijke factoren via het case-controle onderzoek gevolgd door het bestuderen van de correlatie van deze factoren met bepaalde chromosomale afwijkingen binnen de groep van cases. Elke foute informatie kan de studie van dit verband volkomen scheeftrekken. Hier is dus geen sprake van het neutraliseren van de foute antwoorden, en het bekomen van juiste informatie is een onoverkomelijke en prioritaire uitdaging.

De moeilijkheid om correcte antwoorden te bekomen op vragen die op blootstelling betrekking hebben varieert naargelang de aard van de blootstellingsfactor waarop deze vragen betrekking hebben. Het bekomen van correcte antwoorden wordt moeilijk wanneer we precieze informatie wensen over alcoholconsumptie, of over de mate van pesticiden-, solventen- of dioxineblootstelling, of wanneer we bvb zouden vragen "hebt u ooit blootgestaan aan ethyleenoxide?". Nog moeilijker wordt het wanneer we daarenboven een idee willen bekomen

over de periodes in iemands verleden waarin deze blootstellingen plaatsvonden. Nochtans dient de juistheid van de antwoorden op deze laatste soorten vragen de kerninteresse te zijn van de onderzoekers. Het streven naar een uitgesproken standaardisering van de bevraging werd door ons aanzien als een bron van verlies aan precisie, veeleer dan een bron van winst aan vergelijkbaarheid.

Enkele redenen hiervoor zijn:

- de mate waarin men 'gestandaardiseerde' vragen kan aanpassen aan wat men wenst aan de weet te komen is beperkt omdat de vraag altijd begrijpelijk en eenvoudig te beantwoorden moet zijn (ja/nee/weet niet). Dit kan gedeeltelijk opgelost worden door het aantal vragen die verband houden met één enkele blootstellingsfactor te vermenigvuldigen, doch de bevragebaarheid van de geïnterviewde kan daardoor overbelast worden, en het aantal blootstellingsfactoren waarover dient te worden bevraagd moet dan onvermijdelijk worden beperkt. Dit laatste is slechts aanvaardbaar indien er voldoende argumenten zijn om de bevraging tot enkele elementen te beperken.

- de relevantie van de vragen en daaruitvolgende antwoorden wordt wel eens opgeofferd op het altaar van het robot-bevragingsparadigma. Het is bvb. vrij frappant hoe in veel epidemiologisch onderzoek de blootstellingen op het ogenblik van het interview (hoeveel tassen koffie drink je per dag?) rechtstreeks worden ingebracht in de vergelijking en dus de facto worden geprojecteerd naar het verleden. Voor onderzoek naar aandoeningen die zich pas na een zekere latentietijd manifesteren is dit irrelevant wegens mogelijk onjuist: de relevantie van het antwoord kan niet ingeschat worden.

De vraag die ons bezighield, ook vanuit onze eigen ervaring, is op welke wijze elk interview de best mogelijke informatie kon opleveren, waarbij dus aan de juistheid van de bekomen antwoorden absolute prioriteit diende te worden gegeven, en hoe de vergelijkbaarheid daarbij gemaximaliseerd kan worden.

Bij het uitstippelen van de bevragingstrategie werd vertrokken van vier ervaringen:

- de ervaring van een Franse onderzoeksgroep: in dit onderzoek werden 300 MDS cases en controles persoonlijk bevestigd in face to face interviews door een getrainde arts. Daarbij werd gedeeltelijk conversationeel gewerkt (bvb voor de reconstructie van de beroepsloopbaan en de beschrijving van de activiteiten) en daarnaast voor een reeks specifieke blootstellingsfactoren 'in principe' op de bovenvermelde strikt gestandaardiseerde wijze gewerkt aan de hand van een vragenlijst die op zeer directe wijze de agentia vermeldde, waarbij met ja/nee/weet niet diende te worden geantwoord in geval van (huidige of vroegere) blootstelling. Bij de vragen werden semi-kwantitatieve aanduidingen gegeven bvb van minimale blootstellingsduur of werden cumulatieve blootstellingen geschat, doch noch bij de bevraging, noch bij de analyse werd rekening gehouden met de periode 'prior to diagnosis' waarin deze blootstellingen plaatsvonden. Enkele correlaties werden door dit onderzoek in het licht gesteld, doch zijn moeilijk interpreteerbaar.

- de ervaring van een Italiaanse onderzoeksgroep: de recente maar snel groeiende traditie van epidemiologisch onderzoek naar leukemie en andere hematolymfopoiëtische aandoeningen in Italië resulteerde in een gestandaardiseerde vragenlijst die o.a. een beschrijving van de vroegere en huidige beroepsactiviteiten inhield, waarbij niet werd gevraagd naar blootstellingsfactoren. Er werd geopteerd om deze door experts te laten inschatten aan de hand van de beschrijving van de activiteiten. Naast een algemene vragenlijst wordt gebruik gemaakt van een reeks specifieke vragenlijsten per beroep. De bevraging gebeurt hier door getrainde interviewers (verpleegkundigen meestal) in face to face interviews.

- de ervaring van een Zweedse onderzoeksgroep: in een Zweedse case-referent studie werd het verband tussen vermoede risicofactoren en enkele clonale chromosomale aberraties bij MDS en leukemie nagegaan. Dit gebeurde aan de hand van volkstellingsgegevens waarin ook de uitgeoefende beroepen waren opgenomen. In tegenstelling tot wat werd verwacht, aldus de onderzoekers, lijkt het verband tussen deze specifieke chromosomale veranderingen en uitwendige blootstelling niet bevestigd te worden. De onderzoekers wensen deze studie via telefooninterviews uit te diepen.

- de eigen ervaring bij het Europese onderzoek inzake leukemie bij pasgeborenen: in dit case-series onderzoek naar een zeer zeldzame ziekte die haar oorsprong zou vinden in blootstellingen van de moeder tijdens de zwangerschap, werden ouders van pasgeborenen die aan leukemie leden over heel Europa bevestigd in een 'open interview' door twee artsen die een ruime deskundigheid hebben inzake het verkennen van blootstellingsfactoren en het inschatten van blootstellingsgeschiedenissen. Samen met de ouders werden deze blootstellingen op de tijdsschaal van de zwangerschap uitgezet. Daarbij werd gebruikt gemaakt van een vragenlijst die het interview oriënteerde. Deze methode liet toe om een aantal tot nogtoe ongekende blootstellingsfactoren te ontwaren die een oorzakelijke rol kunnen spelen.

Met de buitenlandse onderzoeksgroepen werd uitvoerig overleg gepleegd over hun bevragingstrategie, en hun ervaringen werden getoetst aan onze ervaringen en voorstellen. Daarbij stelden we een centraal uitgangspunt van onze bevragingstrategie voorop dat afwijkt van dat van de buitenlandse onderzoeksgroepen: aangezien over latency bijzonder weinig geweten is, is het noodzakelijk om de blootstellingsgeschiedenissen te reconstrueren en uit te zetten op een tijdsschaal die verscheidene decennia teruggaat, en een vergelijking te maken van blootstellingservaringen in bepaalde periodes voorafgaand aan de diagnose tussen cases en referentiepersonen. Dit houdt meteen in dat een gestandaardiseerde bevraging volstrekt onmogelijk is: het zou absurd zijn om door een reeks vragen waarop met "ja/nee/weet niet" dient te worden geantwoord tot een historische reconstructie te willen komen. Er dient dus een soepele en interactieve bevraging te gebeuren, waarvan het doel echter strak omschreven is. Dit houdt tevens in dat de ondervragers deskundig moeten zijn in de domeinen waarover ondervraagd wordt zoals medisch verleden (ziekten, medicatie, onderzoeken..) of beroepsverleden en blootstellingsfactoren. Deze opties sloten nauw aan bij de evaluatie die de Franse onderzoeksgroep maakte na haar eigen waardevolle aanbreng van nieuwe inzichten: in een verdere fase verwachtte deze groep meer heil van een benadering die een veel kleinere groep cases zou bevragen op een meer diepgaande en beschrijvende wijze waarbij de nadruk moet liggen op het actief construeren van juiste informatie, veeleer dan een strikt gestandaardiseerde bevraging die tot minder juiste antwoorden dreigt te leiden. Van de Italiaanse onderzoeksgroep werd het idee weerhouden om tijdens de ondervraging te streven naar een volledige beschrijving van alle ooit uitgeoefende beroepsbezigheden, zodanig dat in een later fase deze informatie kan vertaald worden in termen van blootstellingsgeschiedenissen, hetzij door de onderzoekers zelf, hetzij door andere experts. Uit de Zweedse ervaring en eigen ervaring besloten we niet zozeer dat het verband tussen bepaalde clonale chromosomale aberraties en uitwendige blootstellingsfactoren een twijfelachtige hypothese is, maar wel dat algemene beroepsomschrijvingen nauwelijks enige waarde hebben voor het inschatten van beroepsmatige blootstellingsgeschiedenissen of voor het inschatten van blootstelling aan bepaalde risicofactoren in het algemeen.

Bevragingmethode

Er werd geopteerd voor een half open interview strategie die als een gestructureerde conversatie verloopt tussen de bevrageerde en de deskundigen inzake de bevrageerde materies, en die moet resulteren in een zo volledig mogelijke beschrijving van activiteiten uit het verleden en de daarmee gepaard gaande reconstructie van mogelijke blootstellingen, waarbij alle gegevens in een reeks strikt omschreven rubrieken op een tijdsschaal worden genoteerd.

De elementen van standaardisering zijn in deze benadering dus niet de vraagstelling of de éénduidigheid van de mogelijke antwoorden, maar wel het samenspel van een geheel van factoren: de ondervragers met hun ervaring, doelstellingen en deskundigheid; de verplichting om te streven naar volledigheid en de wijze van noteren op een tijdsschaal in een éénvormig registratiedossier, dat eigenlijk een vragenlijst is voor de ondervrager, niet voor de ondervraagde.

Uit de eigen ervaring was gebleken dat een dergelijke conversationele wijze van bevragen een aantal belangrijke voordelen kan hebben.

- ❑ het laat de ondervragers de nodige ruimte om de vragen te herhalen en te formuleren in functie van de ondervraagde en aldus te verifiëren of de vraag juist begrepen is en om desnoods corrigerend op te treden;
- ❑ het laat de ondervraagde de nodige ruimte om zich te laten helpen in het reconstrueren van blootstellingsgeschiedenissen, i.p.v. de druk te ervaren die uitgaat van het moeten aanduiden van een voorafbepaald antwoord dat door de ondervraagde vaak als ongenueanceerd wordt ervaren;
- ❑ het laat de ondervraagde de nodige ruimte om de eigen noden of persoonlijke verhalen naar voor te brengen, hetgeen een positief effect heeft op de medewerking en de beleving van het interview;
- ❑ het laat de ondervragers toe om inconsistenties in de antwoorden te verkennen en te corrigeren, in plaats van ze louter te noteren en verder onontgonnen te laten;
- ❑ het laat de ondervragers toe om belangrijke elementen van blootstelling op te sporen die in een gestandaardiseerde bevraging niet aan bod zouden gekomen zijn

Het moge overigens duidelijk zijn dat het aantal te bevragen personen dat nodig is om statistisch relevantie aan te tonen voor een reëel effect lager is naarmate de juistheid van de verzamelde informatie per persoon toeneemt.

Het registratiedocument of 'questionnaire'

Met het registratiedocument (**annex V en VI**) werd op de eerste plaats beoogd om te vermijden dat mogelijks relevante blootstellingen verborgen zouden blijven. Dit werd nagestreefd door het uitvoerig uitspitten van mogelijke blootstelling veroorzakende situaties onder verschillende rubrieken, waarbij de blootstellingsfactoren en de inschattingen van de mate van blootstelling achteraf kunnen gereconstrueerd worden aan de hand van de genoteerde gegevens. Enkele voorbeelden:

De opeenvolgende beroepen van betrokkene, de arbeidstaken, het soort producten waarmee wordt gewerkt worden genoteerd. De 'secundaire' blootstellingen (niet het directe gevolg van de arbeidstaak) worden eveneens genoteerd. Dit kan bvb gaan om reizen naar tropische landen, om blootstelling door agentia van andere afdelingen in hetzelfde bedrijf, om de aanwezigheid van vervuilende industrieën in de nabije omgeving, om het uitvoeren van renovatiewerken in het bedrijf, om blootstellingen in bedrijven waar tijdelijk andere beroepsactiviteiten door derden worden uitgeoefend, etc...

De opeenvolgende woonplaatsen worden genoteerd, en elke woonomgeving in kaart gebracht m.b.t. omgevingsrisicofactoren zoals nabijheid van vliegveld, verbrandingsoven, hospitaal, etc..

De familiestructuur, ziektegeschiedenissen van naastbestaanden worden bevroegd om bvb indirecte blootstelling aan kankerchemotherapie te kunnen vinden;

De individuele ziektegeschiedenis en het medicamentengebruik en bijzondere medische onderzoeken worden afzonderlijk bevestigd en vergeleken. Door het vergelijken van deze elementen kan een redelijk volledige en juiste reconstructie op een tijdschaal worden gemaakt.

Het registratiedocument beoogt daarnaast een semikwantificering van een aantal blootstellingselementen, en dit eveneens op een tijdschaal. Typische voorbeelden hiervan zijn het in kaart brengen van het rokersverleden, het koffieverbruik, het medicamentengebruik.

Enkele praktische tekortkomingen van het registratiedocument zijn de onvoldoende strakke aflijning van enkele begrippen en tussen enkele rubrieken, zoals respectievelijk de definitie van 'vuur' m.b.t. risico voor dioxineblootstelling en de onduidelijkheid inzake de blootstellingen die dienen vermeld te worden onder 'andere blootstellingen aan pesticiden'. Dergelijke tekortkomingen kunnen in een latere fase van het onderzoek worden weggewerkt, en vormen inmiddels geen belemmering voor het samenbrengen van relevante blootstellingsgegevens voor de analyse.

Het objectief dat wordt beoogd met het op deze wijze verzamelen van gegevens is: de vergelijking van de blootstellingsgeschiedenissen van de cases en de controles voor verschillende hypothesen van latency. Een ruwe semi-kwantitatieve blootstellingsinschatting wordt in de vergelijking ingebracht voor een bepaalde factor en voor elk van de vier decennia voorafgaand aan de diagnose.

Ter illustratie van de wijze waarop van de geregistreerde gegevens wordt overgegaan naar blootstellingsevaluatie en de vergelijking tussen cases en controles vermelden we vooreerst het voorbeeld van de pesticiden. De rubriek 'pesticiden' van alle registratiedocumenten wordt aangevuld met gegevens uit andere rubrieken. Vervolgens worden de geschiedenissen van alle cases en controles door deskundigen (in dit geval gebeurde dat voorlopig door de onderzoekers zelf) gerangschikt in een benaderende volgorde volgens de globale intensiteit van blootstelling. Voor het bepalen van de volgorde werd voor elk individu het decennium met de meest uitgesproken blootstelling gebruikt. Op basis van deze rangschikking bekomt men een semi-kwantitatieve schaal die per definitie waardevol is voor de gegevens waarvoor ze moet dienen. Vervolgens wordt voor elk individu per decennium een score gegeven uit deze schaal. In de analyse kan men beslissen over het groeperen van scores en over cut-off points.

Als tweede voorbeeld vermelden we de Ioniserende stralen. Dit betreft zowel medische beeldvorming (RX-opnamen, Tomografie, Scanning, Scintigrafie) als beroepsgebonden blootstellingen, als incidentele blootstellingen. Het objectief is het reconstrueren van de stralingsbelasting per decennium door het samenbrengen van alle geregistreerde gegevens, en het heropbouwen van de meest waarschijnlijke geschiedenis aan de hand van deze gegevens. Voor de medische beeldvorming worden de rechtstreeks bekomen gegevens vergeleken met de ziektegeschiedenis en het medicamentengebruik, en de beroepsrisico's waarvoor periodieke RX Thorax voorgeschreven is, en van daaruit verder bevestigd om het aantal onderzoeken zo volledig mogelijk te reconstrueren. Ook de beschrijving van de onderzoeken door de bevestigde

kan leerzaam zijn om de precieze aard ervan te bepalen. Daarnaast worden professionele blootstellingen ingebracht, zoals het controleren van lasnaden, of de bediening van toestellen voor medische beeldvorming. Zo bekomt men een vollediger beeld van de mogelijke bronnen van ioniserende straling. Door studie van de literatuur en persoonlijke contacten met allerlei deskundigen wordt getracht om tot een ruwe inschatting van de beenmergdosis te komen voor de verschillende onderzoeken of soorten blootstelling, hetgeen moet toelaten om een wegingsfactor toe te passen op verschillende blootstellingssituaties, en zodoende tot een inschatting van de totaaldosis te komen per decennium voorafgaand aan de diagnose voor elke patiënt.

Een derde en nog complexer voorbeeld is de blootstelling aan organische solventen. Hier dienen opnieuw, net als voor de pesticiden, gegevens uit andere rubrieken te worden afgeleid en te worden samengebracht in de solventenrubriek en vervolgens dient een inschatting te worden gemaakt van de globale blootstelling per decennium voor elk individu aan de hand van een uit de antwoorden zelf afgeleide blootstellingsschaal. Hier is de situatie ontzettend complex wegens de grote variabiliteit in mogelijke blootstellingssituaties in professionele en niet-professionele context. Hier werd geopteerd om twee elementen van blootstelling te bevragen: aard van de solventen en aard van de activiteit die tot blootstelling aanleiding kon geven. Zo werden renovatiewerken thuis of in de werkplaats zorgvuldig op een tijdschaal genoteerd, alsook beroepen en hobby's die tot blootstelling aanleiding zouden kunnen geven. Slechts voor enkele gekende solventen of soorten producten werd rechtstreeks naar het gebruik of de blootstelling gevraagd. Door op deze wijze zo nauwgezet als mogelijk een blootstellingsreconstructie te doen, gevolgd door expertinterpretatie, kan men tot een schaalverdeling en een indeling besluiten. Vanzelfsprekend zou deze indeling voor enkele individuen in een bepaalde periode ietwat onnauwkeurig kunnen zijn, doch men kan op deze wijze voor het geheel van de ondervraagde personen tot een zeer betrouwbare opsplitsing komen tussen personen met hoogstwaarschijnlijk lage en hoogstwaarschijnlijk hoge blootstelling, met mogelijks een categorie er tussenin, hetgeen afhangt van de interpretatie van de experts.

Case definitie

Deze staat beschreven in het onderzoeksprotocol (bijlage IV). De cases werden uit een beperkt aantal zeer grote ziekenhuizen gerecruteerd. De redenen hiervoor zijn o.a.: de intensieve contacten die vereist zijn met de diensten voor hematologie, teneinde de patiënten in de meest verantwoorde omstandigheden te kunnen ondervragen en de mogelijkheid om controlepatiënten te recrutereren in hetzelfde ziekenhuis. De intensieve contacten hebben als bijkomend voordeel dat weinig cases van deze zeer ernstige en meestal op relatief korte tijd fatale aandoening verloren gaan voor de bevraging.

Control-definitie

Als controles worden personen genomen die lijden aan drie verschillende aandoeningen: gastro-duodenale ulcera, niertumoren en colonkankers. De idee om zieken als controles te nemen (case-case studies) werd reeds toegepast en theoretisch onderbouwd. Bij de selectie van controles dient een verscheidenheid van aandoeningen in aanmerking te komen, om te vermijden dat een aantal blootstellingsfactoren die mogelijks met zowel de bestudeerde aandoening als de controle-aandoening te maken hebben ongeïdentificeerd zouden blijven.

De controlepersonen werden in hetzelfde ziekenhuis gerecruteerd als de cases waarmee ze voor leeftijd en geslacht gematchd werden. Omdat de controlepersonen in hetzelfde ziekenhuis gerecruteerd worden, en omdat de ziekenhuizen die in aanmerking komen zéér grote referentieziekenhuizen zijn, en de pathologieën bovendien verscheiden zijn, werd beoogd om te voorkomen dat de controlepersonen globaal genomen uit verschillende catchment populaties zouden komen. Deze zijn voor verschillende pathologieën en verschillende ziekenhuizen immers ongelijk, hetgeen een bron van bias had kunnen zijn.

Anderzijds dient de leeftijdsdistributie van de aandoeningen dusdanig te zijn dat zij toelaat om voldoende potentiële controlepersonen te recruteren uit hetzelfde ziekenhuis, zoals aangegeven in annex IV. Uit onze ervaring is gebleken dat er geen overschot was aan potentiële controlepersonen in de zeer grote ziekenhuizen, hetgeen inhoudt dat de insluiting van kleinere ziekenhuizen een probleem geweest ware.

Voor verdere uitleg verwijzen wij naar **annex IV**.

B/IV. RESULTATEN

Het analyseverslag vormt **annex VII**.

We wensen aan dit verslag toe te voegen dat vrijwel alle aangezochte patiënten hun medewerking hebben verleend. Zowel bij cases als bij referentiepersonen was de wil om bij te dragen aan de kennis van de oorzaken en de voorkoming van ziekten zeer groot.

Hieronder volgen enkele bijkomende gegevens.

Controlepersonen:

Voor 14 gevallen werd de eerste controle patiënt van de lijst ondervraagd

Voor 8 gevallen werd de tweede controle patiënt van de lijst ondervraagd

Voor 3 gevallen werd de derde controle patiënt van de lijst ondervraagd

Voor 3 gevallen de vierde controle patiënt

Voor 2 gevallen de zesde controle patiënt

Controle 1	Controle 2	Controle 3	Controle 4	Controle 6
14	8	3	3	2

16 'eerste' controles konden niet geïnterviewd worden:

- 8 overleden
- de huisarts of behandelend ziekenhuisarts adviseerde de patiënt niet te contacteren omwille van:
 - terminaal stadium (2)
 - communicatie probleem (2)
 - erfelijk bepaalde kanker (1)
 - psychologische tegenindicatie (2)
- 1 patiënt werd gecontacteerd maar voelde zich te ziek voor een interview

8 'tweede' controles konden niet geïnterviewd worden

- 4 overleden
- de huisarts of behandelend ziekenhuisarts adviseerde de patiënt niet te contacteren omwille van:
 - communicatieprobleem (2)
 - onbekende reden (1)
 - bij contactname van 1 patiënt werd door de echtgenote een communicatieprobleem gemeld (afasie)

5 'derde' controles konden niet geïnterviewd worden

- 3 overleden
- 1 patiënt werd gecontacteerd maar voelde zich te ziek voor een interview
- bij contactname van 1 patiënt werd door de zoon geweigerd (voor zijn moeder) omwille van psychologische tegenindicatie

2 'vierde' controles konden niet geïnterviewd worden wegens overlijden

2 'vijfde' controles konden niet geïnterviewd worden

- overleden
- weigering van de patiënt

B/V. BESPREKING

Analyses werden reeds verricht voor de meeste rubrieken van de vragenlijst.

We geven hieronder enkele beschouwingen bij de bevindingen uit het analyseverslag. We citeren rubrieken betreffende life-style, medische therapieën, en omgevingsblootstelling door leefmilieu en/of professionele factoren.

Tabaksverbruik:

De resultaten inzake tabaksverbruik illustreren het voordeel dat een historische reconstructie van blootstellingsgegevens en de vergelijking van tijdsperioden 'previous to diagnosis', kan hebben op de herleiding tot actuele rookgewoonten of op een ever/never indeling of zelfs op een inschatting van de cumulatieve dosis.

In de zorgvuldige Franse studie van Dr. Nisse et al met 2 x 203 ondervraagden werd ever/never roken significant in verband gebracht met risico op MDS, doch een dose-respons relatie confirmeerde deze bevinding niet. Dit is eigenlijk een onduidelijk antwoord.

De voorliggende studie waarin slechts 2 x 30 personen ondervraagd werden toonde een statistisch significant verband met rookgewoonten in het vierde decennium voorafgaand aan de diagnose, alsook een aanzienlijk doch niet significant verschil in het derde decennium voorafgaand aan de diagnose, doch geen verband voor recentere periodes. Het zorgvuldig gedetailleerd reconstrueren van de individuele rookgewoonten en het respecteren van de 'nuances' in de analyse geven dus een ondubbelzinnig antwoord op de onderzoeksvraag: een verband met roken in een periode die de diagnose met minstens twee decennia voorafgaat is aangetoond.

Koffieconsumptie:

Koffieconsumptie werd zorgvuldig in kaart gebracht op dezelfde tijdsschaal, en toonde geen statistische significantie, wel een consistent ietwat hoger gebruik bij cases. Er dient aan herinnerd dat met 2 x 30 ondervraagden slechts zéér uitgesproken relatieve risico's op statistisch significante wijze kunnen aangetoond worden. De zorgvuldige bevragsingsmethode en de indeling op een tijdsschaal laat een trend zien alsook consistentie in deze trend, die door een ruwere bevragsingsmethode niet betrouwbaar kan gereproduceerd worden.

Medicamentengebruik:

De bekomen resultaten zijn een indirecte illustratie van de zeer hoge sensitiviteit van de gevolgde onderzoeksmethode. Uit de resultaten blijkt immers duidelijk dat de pooling van medicatie waarvan het mogelijke effect op de bloedvormende organen gekend is, een statistisch significant verband toont met case status. Dit is een uitzonderlijk scherp resultaat voor een studie met slechts 30 matched pairs.

De resultaten voor de afzonderlijke soorten medicatie tonen een consistent beeld van de verwachte correlaties en de verwachte niet-correlaties.

Een merkwaardige vaststelling is de wat verrassende associatie (met deze term wordt louter statistische en dus niet noodzakelijk causale correlatie bedoeld) met antibioticagebruik. Deze bevinding zal het voorwerp uitmaken van verdere analyse. Er dient benadrukt dat vooral majeur antibioticagebruik werd genoteerd. Bij lagere doses kunnen recall fenomenen sterk interfereren met de juistheid van de informatie. Het viel ons op, zowel in deze als in een vroeger verrichte case-series studie omtrent leukemie van pasgeborenen, dat antibioticagebruik moeilijk te

bevragen is, wanneer het te maken had met een veel voorkomende ziekte van voorbijgaande aard, zoals griep of zogenaamde lichte beginnende bronchitis. Deze ziekten worden moeilijker herinnerd. Andere medicaties zoals tegen jicht, maagzweer, cholesterol of hartritmestoornissen worden beter herinnerd omdat ze te maken hebben met ziekten die niet iedereen doormaakt of minder vluchtig zijn dan een beginnende griep met een zgn begin van bronchitis waarvoor soms wel en soms niet antibiotica worden voorgeschreven en vaak weinig disciplinair genomen. Het is onwaarschijnlijk dat dit aan interviewer bias te wijten is. Zulke bias had het gevolg kunnen zijn van het op een selectieve wijze minder noteren van antibioticagebruik voor cases dan voor controlepersonen omdat de interviewers in een eerste fase vooral cases en in een tweede fase vooral controles hebben ondervraagd, en omdat de aandacht in een tweede fase scherper kon zijn voor de medicaties die degelijk kunnen bevraagd worden, en minder scherp voor medicaties waarvan men de bevraging als onbetrouwbaar beschouwt. Elk aanzienlijk antibioticagebruik werd echter genoteerd en enkel dat in aanmerking genomen bij de analyse. In elk geval laat het beperkte aantal cases en controles en de uitstekende contacten met de behandelende artsen en de ondervraagden toe om deze bevindingen te verifiëren.

Blootstelling aan ioniserende straling als gevolg van RX onderzoeken

In afwachting dat meer volledige gegevens over en/of bijkomende 'expert judgement' voor de mogelijke beroepsmatige blootstelling aan ioniserende stralingen van enkele individuen kan bekomen worden, werd reeds een analyse gemaakt van de mogelijke correlatie met ondergane medische RX onderzoeken. Hier werd geen verband aangetoond.

Pesticidenblootstelling

De blootstellingsgegevens aan pesticiden werden zo volledig als mogelijk verwerkt, en zoals aangegeven in bijlage VII gekwalificeerd en gecodeerd. De lectuur van deze bijlage moge volstaan om de methode van 'expert judgement' te illustreren. Het resultaat is een statistisch significante correlatie. Ook deze bevinding is merkwaardig omwille van het beperkte aantal cases en controlepersonen, en is eens te meer een argument voor de gevoeligheid van de gebruikte onderzoeksmethode.

Solventenblootstelling

Alle gegevens die bij de bevraging werden bekomen en die een rechte of onrechte aanduiding konden vormen voor blootstelling aan solventen werden samengebracht: beroepsmatige blootstelling, blootstelling als gevolg van renovatiewerken of hobby's, rechtstreeks gerapporteerde blootstelling in antwoord op vragen naar specifieke producten, etc... Op basis daarvan werd een indeling gemaakt in blootstellingscategorieën, en werd een overeenkomstige score toegewezen aan alle cases en controles. Er werd een niet statistisch significant effect vastgesteld. Aangezien er beperkte epidemiologische aanduidingen zijn voor een mogelijk verband tussen niertumoren en blootstelling aan solventen, werd een bijkomende analyse gedaan met uitsluiting van de matched pairs met niertumor als controle. De power van een dergelijke vergelijking verkleint meteen. Er werd een verschuiving van de p-value vastgesteld naar een lagere eveneens niet significante waarde. Deze weliswaar niet significante bevindingen zijn echter wel consistent met deze van de studie van het CHU te Lille.

Dioxineblootstelling

Er werd in het onderzoek bevestigd op een reeks elementen die moeten toelaten om het dioxineblootstellingsverleden te reconstrueren. Men mag aannemen dat terwijl biologisch onderzoek daar duidelijke aanduidingen van zou kunnen geven, een zelfs zorgvuldige bevraging door deskundigen maar een beperkte gevoeligheid heeft. Een volledige inschatting van een mogelijke dioxineblootstellingsverleden op grond van de bekomen gegevens gebeurde nog niet. Er werd nog geen inschatting gedaan van de mogelijke dioxineblootstelling als gevolg van de beroepsactiviteiten van een aantal cases en referentiepersonen. Contact met vuurhaarden en verbranding van PVC werd zorgvuldig bevestigd, doch gaf nauwelijks betekenisvolle aanduidingen. Ook de afstand van woonplaats en arbeidsplaats tot huisvuil- en industriële

verbrandingsovens en ziekenhuizen (die veelal een oven hebben waarin PVC producten worden verbrand) werd systematisch nagegaan. Enkel voor deze laatste elementen werd een analyse gedaan. De controles waren significant meer blootgesteld dan de cases.

Verband tussen clonale chromosomale veranderingen en pesticidenblootstelling, medicamentengebruik en solventenblootstelling.

Een eerste exploratie van dit verband toonde een positieve trend voor pesticidenblootstelling en het voorkomen van clonale wijzigingen van de chromosomen 5, 7 en 11, die in genetische kringen worden aangezien als handtekeningen van uitwendige blootstelling. Deze associatie is het sterkst voor de meest recente decennia.

Voor medicamentengebruik werden de geschiedenissen niet voor de afzonderlijke periodes geanalyseerd. Het verband werd bestudeerd aan de hand van gepoolde gegevens waarbij door de deskundigen wel een zwaarder gewicht werd gegeven aan de meer recente decennia. Er werd geen statistisch significant verband vastgesteld.

Voor solventenblootstelling, dat in tegenstelling tot de twee voorgaande indelingen niet op statistische significantie wees in de case-referent studie werd een niet statistisch significant omgekeerd verband vastgesteld.

Opvallend in deze bevindingen is dat één associatie die duidelijk significant was in de case-control studie eveneens een duidelijke associatie laat zien tussen clonale chromosomale anomalieën 5, 7 en 11 en de mate van blootstelling van de cases enerzijds, en dat een niet significante associatie in de case-control studie géén verband laat zien. Deze beperkte bevindingen zijn consistent met de ervaring uit de kliniek dat de chromosomale anomalieën 5, 7 en 11 met uitwendige blootstelling geassocieerd zijn. Deze bevindingen dienen het voorwerp uit te maken van verdere verfijnde benadering en onderzoek op grotere aantallen.

Deze voorlopige bevindingen laten in elk geval toe om te veronderstellen dat de gegevens die in dit onderzoek werden verzameld zouden kunnen bijdragen tot een verdere verkenning van het thema van het onderzoeksopzet: 'blootstellings specifieke afwijkingen van het genoom'.

Professionele factoren die geen deel uitmaken van de bovengenoemde:

Voor alle cases en controlepersonen werd een zo goed mogelijk reconstructie gemaakt van het beroepsverleden. In vele gevallen waren de ondervraagden niet in staat om uitsluitel te geven over een aantal beroepsgebonden blootstellingsfactoren. Via verschillende kanalen worden de inspanningen verdergezet om hierover meer precieze informatie aan de weet te komen. Een aantal analyses kunnen niet afgewerkt worden zolang deze informatie niet volledig is verzameld voor alle cases en controles, voor enkele blootstellingsfactoren zoals zware metalen of dioxines.

De zogenaamde professionele factoren zullen zoals dat het geval was voor pesticiden en solventen samengebracht worden met blootstellingen aan dezelfde factoren buiten de beroepssfeer voor de analyse.

CONCLUSIES

Evaluatie van de bevragingmethode

Uit de resultaten van het onderzoek blijkt dat de gevolgde bevragingmethode een merkelijk hogere gevoeligheid heeft dan de meer klassieke benaderingen. Het onderzoek bevestigt de hoger aangegeven verwachtingen t.a.v. de ontwikkelde conversationele bevragingmethode. Daarnaast blijkt de methode toe te laten om:

- bij gelijke omvang van het aantal ondervraagden met grotere zekerheid risicofactoren te kunnen identificeren door case-controle onderzoek
- tot even significante conclusies te komen door onderzoek op kleinere groepen personen
- meer betrouwbare antwoorden en dus meer betrouwbare conclusies te bekomen¹
- oorzakelijk epidemiologisch onderzoek te verrichten voor minder frequente aandoeningen
- correlaties tussen bepaalde chromosomale afwijkingen en bepaalde blootstellingsfactoren te bestuderen

Blootstellingsspecifieke clonale chromosomale veranderingen bij kwaadaardige aandoeningen.

Het is opvallend dat voor MDS cases een uitgesproken verband werd aangetoond tussen bepaalde clonale chromosomale veranderingen (gekend voor het verband met uitwendige blootstellingsfactoren) en een blootstellingsfactor die als zeer significant uit het case-referent onderzoek naar voren kwam, met name de pesticidenblootstelling. Deze gecombineerde significantie uit éénzelfde onderzoek op een relatief beperkte populatie vormt een duidelijke bevestiging van de zin van verder onderzoek naar blootstellingsspecificiteit in de aantasting van het genoom. Tevens vormt het een nadrukkelijke aanduiding dat de gevolgde epidemiologische methode die door de onderzoeksteams volledig werd uitgewerkt in functie van dit onderzoek

¹ Eén voorbeeld ter illustratie: een nachtwaker antwoordde negatief op de vraag naar pesticidenblootstelling, doch bleek bij zorgvuldige bevraging enkele jaren zeer intens te zijn blootgesteld tijdens zijn nachtrondes. Deze informatie werd bekomen door zorgvuldig converseren over de plaatsen waar hij overal als nachtwaker kwam, en door de kennis van de ondervragers over de mogelijke risico's in die bedrijven. Enkel aldus converserend werd vernomen dat de nachtwakers werden ingezet om in een experimenteel pesticiden-laboratorium alle applicatiepunten nauwgezet te controleren en deze terug in de bodem te stoppen indien nodig. De daartoe aanwezige individuele beschermingsmiddelen bleken in elk geval niet correct te zijn gebruikt. Het is eerder onwaarschijnlijk dat een dergelijk verhaal aan bod zou gekomen zijn in een gestandaardiseerde bevraging. Refererend naar de bovengeciteerde studies kan dit geïllustreerd worden: in de Franse studie werd in een halfopen bevraging geïnformeerd naar professioneel verleden en taken. Tot zolang wij de patiënt op die wijze ondervroegen, werd op geen enkele wijze melding gemaakt van enige blootstelling aan scheikundige stoffen en werd dit zelfs ontkend bij navraag. Enkel het interactief vragen (aandringend converseren) en de deskundigheid van de ondervragers opende de weg naar een mogelijks zeer significante blootstellingsgeschiedenis. Het is ook onwaarschijnlijk dat het meer gestandaardiseerde gedeelte van de Franse vragenlijst een positief antwoord zou opgeleverd hebben voor pesticidenblootstelling. De Italiaanse studie bevat geen specifieke vragenlijst voor het beroep van nachtwaker. Het is zeer onwaarschijnlijk dat de bevraging met de algemene beroepslijst hier een positief antwoord zou opgeleverd hebben. Het hoeft verder geen betoog dat de majeure blootstellingsgeschiedenis onmogelijk aan bod kon komen in de Zweedse studie.

zéér gevoelig is en derhalve een aangewezen weg is voor verder epidemiologisch onderzoek naar blootstellings specifieke clonale veranderingen bij kwaadaardige aandoeningen.

De voorliggende resultaten zijn een aanmoediging om verder te gaan met het onderzoek inzake deze thematiek, en laten verhopend dat uit dit onderzoek elementen kunnen naar voor komen die nuttig kunnen zijn voor het uitwerken van criteria voor de herkenning en erkenning als beroepsziekten van bepaalde kwaadaardige aandoeningen die thans niet als dusdanig worden herkend, alsook voor een beter begrip van de elementen van gezondheidsbewaking ter voorkoming van deze ziekten.

Met het oog op deze laatste doelstelling werd een onderzoeksvoorstel uitgewerkt dat het derde luik vormt van dit verslag.

De wenselijkheid van een MDS Register

Om meer inzicht te verwerven in de relatie tussen uitwendige factoren en het voorkomen van MDS wordt voorgesteld een MDS Register uit te bouwen. Zulk register moet een reeks relevante gegevens bevatten die toelaten om vele facetten van het verband tussen de aandoening en omgevingsfactoren te bestuderen. Dit register moet alle incidente MDS gevallen omvatten die relevant kunnen zijn in verband met milieublootstellingen in België. Een dergelijk register bestaat niet. Een hospital-based registratie is het meest aangewezen vertrekpunt om te trachten volledigheid te verwezenlijken. Patiënten met MDS worden zowel in hematologische als in oncologische diensten gevolgd. Beide diensten moeten dan ook betrokken worden in het op te bouwen netwerk dat alle ziekenhuizen moet omvatten.

Kwaliteitsvolle en volledige registratiegegevens moeten toelaten om deze ernstige ziekte in kaart te brengen. De diagnose is niet eenvoudig en vereist niet enkel een zeer grondig inzicht in hematologische aandoeningen, maar tevens een zeer strak hanteren van gestandaardiseerde diagnostische criteria. De complexe klinische en genetische heterogeniteit van de aandoening, de evoluerende inzichten daaromtrent en de noodzaak van identificerende gegevens om terugkoppeling naar leefomgevingfactoren te kunnen doen, maken een zeer specifieke aanpak van de registratie noodzakelijk, hetgeen vrijwel uitsluitend via een bijzonder register kan worden gerealiseerd.

Het is van essentieel belang dat de gegevens in een MDS register op directe en eenvoudige wijze identificeerbaar zijn, en dus onmiddellijk identificerend kunnen gemaakt worden voor geautoriseerde onderzoekers. Het systeem van de huidige kankerregisters waarbij enkel de notifiërende arts de terugkoppeling kan doen zou haaks staan op de verwezenlijking van de doelstelling die met dit MDS register dient te worden nagestreefd. Zo zal in bepaalde gevallen een individuele bevraging van alle patiënten behorend tot één cluster noodzakelijk zijn om het verband met milieufactoren te kunnen bestuderen. Het register moet ook een centrale registratie van genetische gegevens inhouden en gekoppeld zijn aan een centraal georganiseerde bewaring van beenmerg en bloedstalen. Dergelijke biobank moet toelaten om telkens als het nodig is terug te grijpen naar oorspronkelijk biologisch materiaal in het licht van nieuwe inzichten of nieuwe diagnostische middelen.

Ook onderzoekers in andere Europese landen met bijzondere belangstelling voor epidemiologisch onderzoek van MDS, meer bepaald in Noord Frankrijk en Toscane (zie lijst reeds beschikbare expertise), voelen de noodzaak aan om een specifiek register uit te bouwen. Door een gecoördineerde aanpak zouden de mogelijkheden tot valorisatie aanzienlijk worden vergroot.

C. ONDERZOEK NAAR SUSCEPTIBILITEITSFACTOREN: STUDIEVOORSTEL

Het onderzoeksvoorstel vormt annex VIII bij dit rapport. Het betreft een onderzoek naar de betekenis van reversiebele perifere bloedbeeldafwijkingen bij werknemers blootgesteld aan een gekend myelotoxisch agens (benzeen) voor het risico op MDS/leukemie. Hiertoe wordt voorgesteld om de frequentie van bepaalde genetische susceptibiliteitsfactoren te vergelijken tussen werknemers met bepaalde reversiebele perifere bloedbeeldafwijkingen en patiënten die lijden aan MDS/leukemie. Deze studie zou het antwoord kunnen geven op de vraag naar het belang van bepaalde perifere bloedbeeldafwijkingen als mogelijke indicator van een risico op MDS/leukemie en aldus een beter inzicht verschaffen in de gezondheidsbewakingsstrategie. Daarnaast moet het antwoord op de vraag of genetische susceptibiliteitsfactoren een effect modifier kunnen zijn voor het risico op MDS/leukemie bij blootgestelden aan myelotoxische agentia toelaten om de studie van het verband tussen blootstellingen en specifieke chromosomale anomalieën verder te verfijnen. De studie van het verband tussen de frequentie van deze susceptibiliteitsfactoren en specifieke clonale chromosomale veranderingen bij MDS/leukemie kan op haar beurt het verlengstuk vormen van dit nieuwe onderzoeksvoorstel.

VI. BIJLAGEN

A. MYELOYDYSPLASTISCH SYNDROOM: ANALYSE VAN HISTORISCHE DATABASE

Annex I	Ruwe database
Annex II	Confirmed databases
Annex III	Extended confirmed database

B: MYELOYDYSPLASTISCH SYNDROOM: CASE-REFERENT STUDIE

Annex IV	Onderzoeksprotocol van de case-referent studie
Annex V	Registratiedocument voor de interviews met cases en controles
Annex VI	Vragenlijsten voor specifieke beroepen: enkele voorbeelden
Annex VII	Analyseverslag

C: ONDERZOEK NAAR SUSCEPTIBILITEITSFACTOREN - VOORSTEL

Annex VIII	Voorstel voor een nieuw onderzoek.
------------	------------------------------------

Annex I

MDS - Analysis of historical database: preliminary database

Author:

Dr. Alix Collard

Investigators:

Dr Karel Van Damme

Dr. Ludwine Casteleyn

Dr. Alix Collard

Dr. Lucienne Michaux

Prof. Dr. Anne Hagemeyer

Prof Dr. Herman Van den Berghe

MDS: Analysis of historical database

Results of the exploratory analysis of the preliminary MDS database

Introduction

From the existing literature (Third MIC Co-operative group, 1988, Aul et al, 1995, Rodella et al, 1993), the hypothesis was raised that in a number of MDS cases, a genomic lesion might be linked with a specific exposure of the patient. The retrospective study which is the subject of this preliminary report has been realised in close collaboration with the Center for Human Genetics (Prof. H. Van den Berghe), UIA (Dr. L. Casteleyn and K. Van Damme) and FUNDP (A. Collard).

This report concerns 4134 patients who have been referred to the Center for Human Genetics (Leuven) between 1977 and 1995, and who give rise to 8274 observations. These observations were selected from the root database of the centre according to the following inclusion/exclusion criteria:

- tissue type 500, 501 or 505
- sample date after 01/01/75 and before 01/01/97
- diagnostic containing one of the following character sequences (where a * might represent any series of characters and ? represents any character):

MDS	*RAEB*
SMD	*R.A.E.B*
M.D.S	*CMM*
S.M.D	*R.A*
D?SPL	*REFRAC*
D?SMYELOP	*CYTOPENIE*
D?SERYTHROP	*AREB*
*RA *	*LEU?OPENIE*
RAE	*THROMBOCYTOPENIE*
PANCYTOPENIE	*A.R.E.B*
C.M.M	*LMMC*
L.M.M.C	*ACHRES*
ACRES	*A.S.I.A*
A.I.S.A	*C/*

These different strings correspond to different possible qualifications of diagnostic compatible with MDS, except for " C/ " which is a qualificator given for patients with confirmed MDS.

The selection of the patients from the extracted database aimed at excluding patients with diagnosis probably different from MDS. All patients with only one sample date, a normal karyotype (as indicated by a status code different from 25) and a diagnosis similar to thrombopenie, leucopenie, granulopenie or dyserythropenie were excluded as were some patients with a confirmed diagnosis (indicated by a slash) which was judged too different from MDS.

The information collected on these patients and relevant for this study is described in 16 variables which are summarised in table 1.

Table 1: Variable list for the MDS retrospective study database.

Variable name	Variable content
Patid	Patient identification number
Sex	Sex
Zip	Zip code (residence)
Age	Age
birth.date	Birth date
sample.number	Sample number
sample.date	Sample date
inst*	Hospital
tissue.1	Tissue type. first sample
tissue.2	Tissue type . second sample
Diag	Diagnostic
Lab	Lab number
Clone	Clone number
Number.cell	number of cells within one clone
Karyo	Karyotype
status*	status code

* see appendix 1 for codes

A 17-th variable was also available, indicating if the patient had a 'confirmed' diagnosis or not.

Data quality

Concerning the demographic variables, some errors were detected especially concerning the variable age as given by the database variable. Indeed, some patients were specified as having a negative age. This is due to the incomplete coding of the birth year, which raises problems for people who were born before 1900. Corrections have been made in the database for all patients with negative ages. Moreover, all checks have been made to check for all young (less than 20 years old) patients in order to be sure that there wasn't a similar problem for patients with low ages.

Results

Demographics

The whole population considered in this report comprises 2 172 males (52.5%) for 1 946 females (47.1%), 12 patients with code U(nkwon), and 4 patients with no information available for sex. The sex ratio is estimated to be 1.1 and significantly different from 1 (p-value = 0.0002). This finding is similar to the result of Dr. C. Nisse who found a preponderance of male patients (62%) (Nisse, 1996), as is usually admitted (Aul et al, 1995).

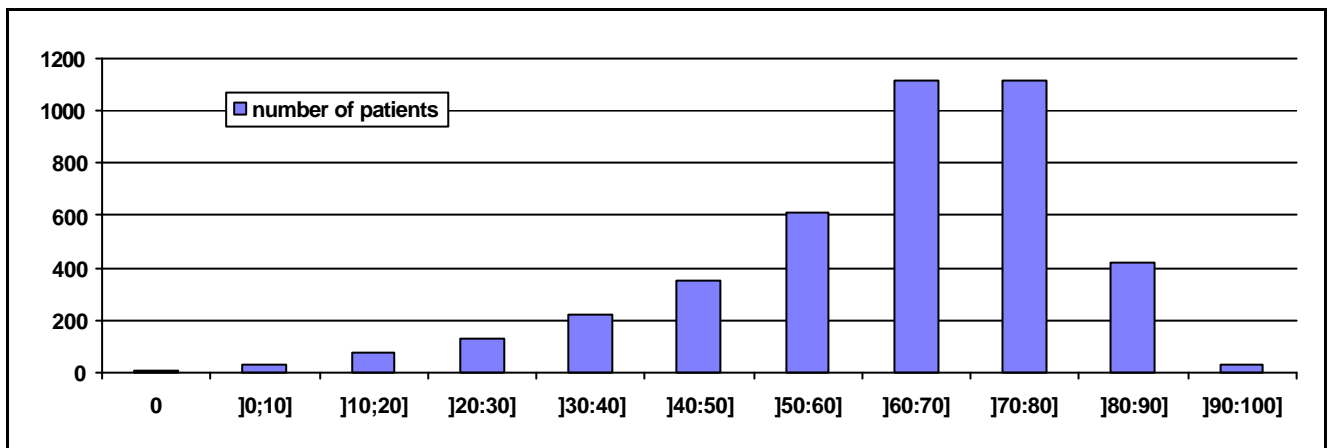
The age distribution for the study population is summarised in table 2 and a histogram is given in figure 1. Seven hundreds and eighty five (785, 19.0%) are less than 50 years old at inclusion in the database, and 1847 (44.7%) are less than 65 years old. This age distribution seems to show slightly younger patients than usually reported. For example, Dr. C. Nisse (Nisse, 1996) reported a mean age of 67.2 ± 11 years (p-value = 0.0003). Another study (Third MIC Co-operative study group, 1987) reported a median age of approximately 70 years and another reported a median age of 73 years with only 11 patients among a cohort of 141 under 60 years of age (Mufti et al, 1985). It is usually admitted that more than 75% of the

patients are older than 60 years at diagnosis (Aul et al, 1995). In our own population, 2772 (67%) patients only are more than 60 years old. Nevertheless, a median age as low as 58 years was already reported in a preceding study (Morra et al. 1990).

Table 2: Summary of age distribution for population database.

Number of patients	4134
Number of NA	22
Mean	62.89
Standard deviation	16.89
Minimum	0
Median	66
Maximum	99

Figure 1: Age distribution for patients in the global database.



Hospital distribution.

The number of patients included by each of the 19 hospitals included in this study is given in table 3, together with the number of patients entered by year.

Table 3: Distribution of patients by year of entry and hospital. For clarity, hospitals with less than 5 patients for the whole study period are not included in the table

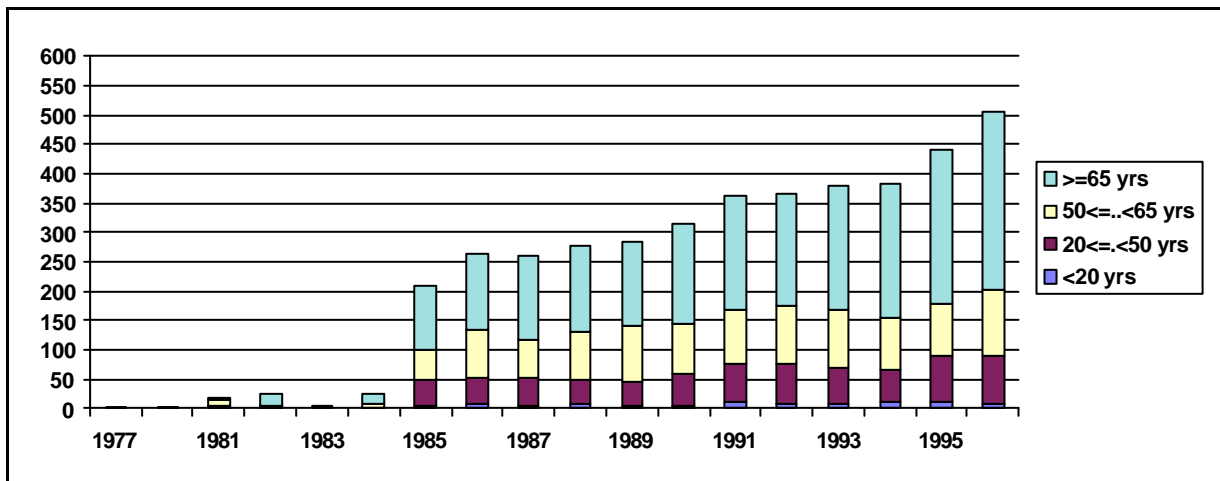
Year	A	B	G	J	L	LS	LX	MG	N	P	R	RG	RH	RP	totals
1977	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
1980	0	0	0	0	0	0	0	0	0	1	0	0	1	0	2
1981	0	1	1	0	2	0	0	2	0	2	1	0	4	0	18
1982	0	1	4	0	3	0	0	7	0	2	1	0	2	0	24
1983	0	1	0	0	2	0	0	0	0	0	0	0	0	0	3
1984	1	1	0	0	5	0	0	3	2	2	0	0	9	0	26
1985	3	28	18	0	42	0	0	26	2	23	0	4	63	0	210
1986	3	36	23	0	67	1	0	41	1	39	0	4	48	1	269
1987	7	42	12	0	48	2	0	29	1	63	0	2	52	1	263
1988	6	37	9	0	41	1	0	29	2	63	1	9	74	3	280
1989	4	34	1	11	52	3	0	30	2	54	0	7	83	2	286
1990	9	39	0	25	55	0	6	26	1	54	5	11	84	0	319
1991	15	50	0	23	63	0	0	42	0	55	11	4	93	3	361
1992	15	50	0	27	57	0	0	41	0	51	26	10	84	3	366
1993	15	58	0	14	36	0	0	40	0	107	17	11	78	0	378
1994	12	64	0	22	31	0	0	39	0	115	19	5	68	6	383
1995	20	68	0	20	37	0	0	39	0	153	22	2	71	6	439
1996	13	70	0	43	19	0	0	41	0	199	25	0	80	4	506
Totals	123	580	68	185	560	7	6	435	11	983	128	69	904	29	4134

* there are 3 patients with unknown institution

Time evolution

The number of patients newly entered each year in the database has been continuously increasing along the 19 years of inclusion covered in this report (see figure 2).

Figure 2: number of patients entered each year in the database. Different grey levels correspond to different age categories. See legend for explanation.



Diagnosis

Among the 4 134 patients and 8274 observations in the database, 3 278 different strings of characters were used to qualify the patients diagnosis. This number essentially reflects the absence of standardisation for this variable in the present database. This will have to be corrected in the forwarded database development to allow for a more interesting use of this information in the future.

Karyotypes.

The karyotypes distribution according to the database variable *status* is given in table 4.

Table 4: Karyotypes distribution according to the database variable *status*.

Status code	1	3	8	10	15	16	19	25	35	45
Karyotype category					Normal		Niet Afgewerkt	Abnorma I	No mitose	
Number of patients	2	4	1	5	1914	6	2	804	1009	386
Percentage of total population			< 0.1		38.8	0.1	< 0.1	19.4	24.4	9.3

The karyotype variable *karyo* takes 1076 different values for the 8274 observations contained in the database. Among the 4134 patients, 973 (23.5%) have at least one observation different from NA, 46,XX and 46,XY. The age and sex distribution for these patients is summarised in table 5.

Table 5: Summary of age distribution for patients with at least one karyotype problem.

Number of patients	973
Sex distribution: F	392
M	580
U	1 (p-value < 0.0001: sex ratio different from 1)
Number of NA	3
Mean	64.30
Standard deviation	16.63
Minimum	0
Median	68
Maximum	95

Appendix 1: Database codes

Variable Hospitals

A:	Antwerpen	MG:	Mont-Godinne
B:	Brugge	N:	Namur
C:	Charleroi	P:	Périphérie
G:	Gent	R:	Gasthuisberg
J:	Jolimont	RG:	Gasthuisberg-Oncology
L:	St Luc	RH:	Gasthuisberg-Hemato
LS:	St Luc	RP:	Gasthuisberg-Pediatrie
LX:	St Luc		

Variable Tissue.1

500	Blood
501	Rate
505	Bone marrow

Variable Tissue.2

Sample number of the considered tissue as defined by tissue type.1. For example,
500/1: Blood, first sample
500/7: Blood, seventh sample

Variable status

15	Normal karyotype
16	Request for new sample
19	Not performed
25	Abnormal karyotype
35	No mitose
45	Normal karyotype with comment

Annex II

MDS - Analysis of historical database: confirmed database

Author:

Dr. Alix Collard

Investigators:

Dr Karel Van Damme

Dr. Ludwine Casteleyn

Dr. Alix Collard

Dr. Lucienne Michaux

Prof. Dr. Anne Hagemeijer

Prof Dr. Herman Van den Berghe

MDS: analysis of historical database

Results of the exploratory analysis of the MDS database of CONFIRMED patients

Introduction

This report concerns 602 patients who have been referred to the Center for Human Genetics (Leuven) between 1980 and 1995, and who give rise to 1727 observations. These observations were selected from the root database MDS-T extracted from the CME Oracle database according to the following inclusion/exclusion criteria:

- tissue type 500, 501 or 505
- sample date after 01/01/75 and before 01/01/97
- diagnostic containing one of the following character sequences (where a * might represent any series of characters and ? represents any character):

MDS	*RAEB*
SMD	*R.A.E.B*
M.D.S	*CMM*
S.M.D	*R.A*
D?SPL	*REFRAC*
D?SMYELOP	*CYTOPENIE*
D?SERYTHROP	*AREB*
*RA *	*LEU?OPENIE*
RAE	*THROMBOCYTOPENIE*
PANCYTOPENIE	*A.R.E.B*
C.M.M	*LMMC*
L.M.M.C	*ACHRES*
ACRES	*A.S.I.A*
A.I.S.A	*C/*

These different strings correspond to different possible qualifications of diagnostic compatible with MDS, except for « C/ » which is a qualificator given for patients with confirmed myelodysplasie.

The selection was made in order to give more confidence that these selected patients diagnosis were indeed MDS. The selection criteria included the patients with a confirmed myelodysplasia (« C/ » in the diagnosis) plus patients confirmed by clinician. This selection gives little more weight to hospitals who have sent confirmation lists(namely Mont-Godinne, St Luc and Jolimont), but this is not expected to bias the population demographic or genetic structure.

The information collected on these patients and relevant for this study is described in 16 variables which are summarised in table 1.

Table 1: Variable list for the MDS retrospective study database.

Variable name	variable content
Patid	patient identification number
Sex	Sex
Zip	zip code (residence)
Age	Age
birth.date	birth date
Sample.number	sample number
Sample.date	sample date
Inst	Hospital
tissue.1	tissue type. first sample
tissue.2	tissue type. second sample
Diag	Diagnostic
Lab	lab number
Clone	clone number
Number.cell	number of cells within one clone
Karyo	Karyotype
Status	status code

* see appendix 1 for codes

Data quality

Concerning the demographic variables, some errors were detected especially concerning the variable age as given by the database variable. Indeed, some patients were specified as having a negative age. This is due to the incomplete coding of the birth year, which raises problems for people who were born before 1900. Corrections have been made in the database for all patients with negative ages. Moreover, all checks have been made to check birth dates for all patients with coded ages less than 20 years. This didn't induce any further correction.

An anomaly has also been detected for patient 87 782 for whom one observation is dated from 1977, but not classified following the chronological order as usually in the database. This patient observations were sorted by sample date to correct for this irregularity.

Several patients were checked positive by clinician and however not included in the database reported on in this work, because these patients were actually not in the initial root database we are extracting our information from. This inconsistency is still under check and is likely to lead to a revision of the extraction criteria from the ORACLE database.

Results

Demographics

The whole population considered in this report comprises 351 males (58.3%) for 248 females (41.2%), and 3 patients with no information available for sex. The sex ratio is 1.41, significantly different from 1 (p -value < 0.0001). This finding is comparable the result of Dr. C. Nisse who found a preponderance of male patients (62%) (Nisse, 1996), as is usually admitted (Aul et al, 1995).

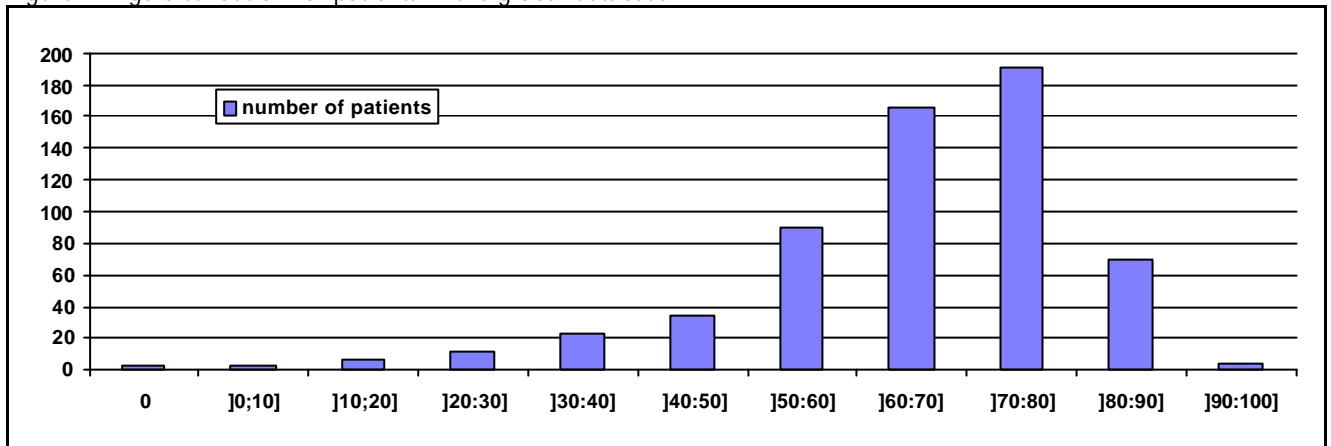
The age distribution for the study population is summarised in table 2 and a histogram is given in figure 1. Seventy four patients (74, 12.3%) are less than 50 years old at inclusion, and 222 (36.9%) are less than 65 years old. This age distribution seems to conform with those usually reported. For example, Dr. C. Nisse (Nisse, 1996) reported a mean age of 67.2 ± 11 years (p -value for two sided comparison of biot sample age characteristics: 0.23). Another study (Third MIC Co-operative study group, 1987) reported a median age of

approximately 70 years and another reported a median age of 73 years with only 11 patients among a cohort of 141 under 60 years of age (Mufti et al, 1985). It is usually admitted that more than 75% of the patients are older than 60 years at diagnosis (Aul et al, 1995). In our own population, 440 (73.1%) patients are more than 60 years old.

Table 2: Summary of age distribution for population database.

Number of patients	602
Number of NA	4
Mean	65.83
Standard deviation	15.09
Minimum	0
Median	69
Maximum	99

Figure 1: Age distribution for patients in the global database.



Hospital distribution.

The number of patients included by each of the 18 hospitals included in the study is given in table 3, together with the number of patients entered by year. Apart from the peripheral hospitals which account for more than 40% of the total recruitment over the last 5 years, the expected « over representation » of Jolimont, Saint Luc and Mont Godinne is found back and these hospitals account for between 12 and 18 % of the total recruitment of the last 5 years. During 1996, only 11 patients have been recruited from those three hospitals with a total number of 21 inclusion in the database.

Table 3: Distribution of patients by year of entry and hospital. Percentages given in the totals columns correspond to confirmation percentages.

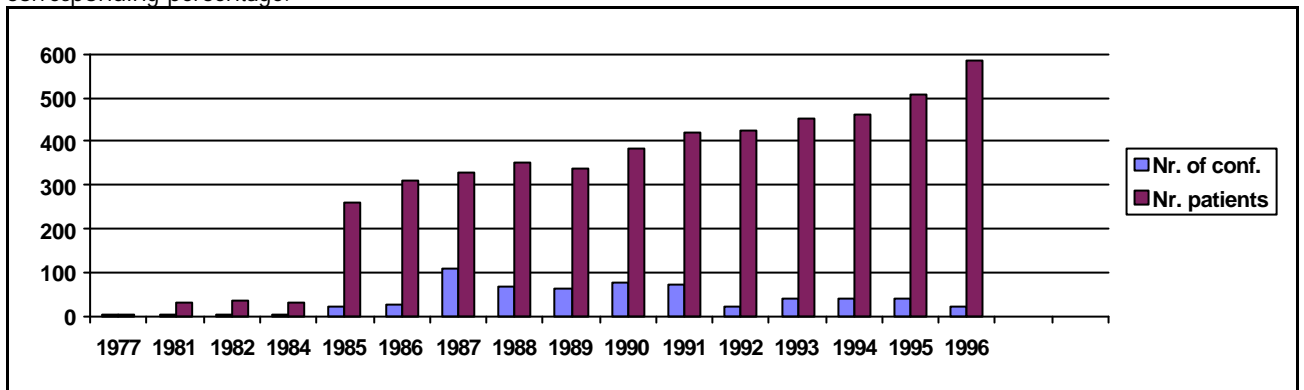
Year	A	B	G	J	L	LS	LX	MG	P	R	RG	RH	RP	totals (%)
1977	0	0	0	0	0	0	0	0	0	0	0	1	0	1 (100)
1981	0	0	0	0	0	0	0	1	0	0	0	0	0	1 (3.4)
1982	0	0	0	0	1	0	0	0	0	0	0	0	0	1 (2.9)
1984	0	0	0	0	3	0	0	0	0	0	0	0	0	3 (10.3)
1985	0	1	1	0	13	0	0	1	0	0	0	4	0	20 (7.7)
1986	0	2	0	0	16	0	0	4	1	0	0	3	0	26 (8.3)
1987	6	11	8	0	31	1	0	13	35	0	0	3	0	108 (32.7)
1988	0	2	4	0	15	0	0	10	33	0	2	1	0	69 (19.5)
1989	1	1	0	5	11	1	0	10	14	0	0	20	1	64 (19.0)
1990	0	1	0	16	10	0	2	6	23	0	0	18	0	76 (19.8)
1991	1	0	0	16	8	0	0	8	15	4	0	17	0	69 (16.3)
1992	0	0	0	5	3	0	0	5	7	0	0	3	0	23 (5.4)
1993	0	0	0	6	8	0	0	8	10	1	0	5	0	38 (8.4)
1994	0	0	0	11	2	0	0	3	15	2	0	4	1	38 (8.2)

1995	0	1	0	4	4	0	0	4	23	1	0	2	0	39 (7.7)
1996	0	0	0	3	2	0	0	6	10	0	0	0	0	21 (3.6)
Totals	8	19	13	66	127	2	2	79	186	8	2	81	2	594*
(%)	(6.2)	(2.8)	(17.3)	(32.3)	(18.4)	(22.2)	(18.2)	(13.8)	(15.9)	(5.3)	(2.6)	(7.7)	(3.9)	(12.2)

* there are 7 patients with unknown institution

The number of patients newly confirmed each year in the database is extremely variable and has been quite low along these last years. This is further illustrated in figure 2. As illustrated by this graphic the percentage of confirmation is rather low. Coming from its maximum in 1987, it has decreased to around 15% during the next three years and is less than 10% since 1992, even if a trend towards improvement might be noted for this period.

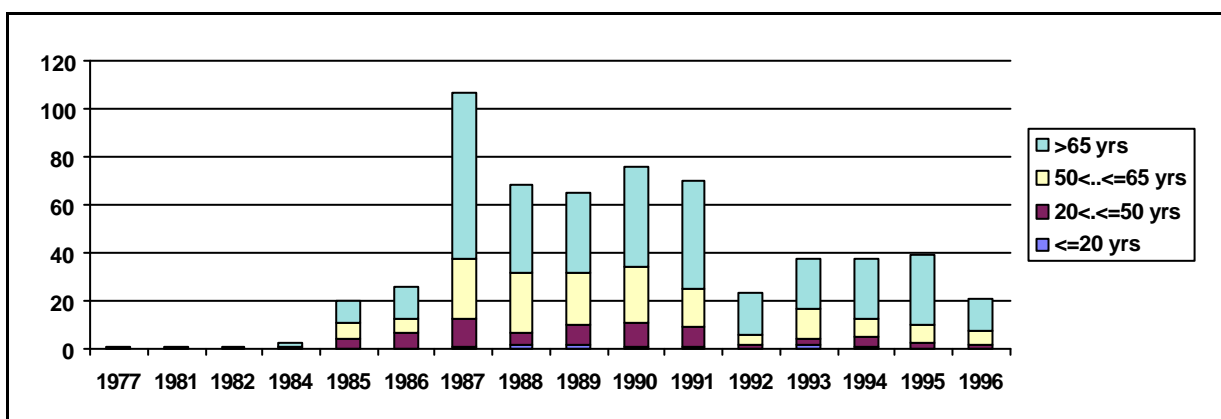
Figure 2: number of patients newly confirmed each year in the database and corresponding percentage.



Time evolution

The number of patients newly entered each year in the database has been continuously increasing along the 16 years of inclusion covered in this report as illustrated in figure 2. However, as already noted, the picture given by the confirmed database is slightly different, probably partly due to low response rate to the confirmation procedure. A more detailed picture of the patients included in the « confirmed » database since 1977 is given in figure 3, together with some details on the age repartition.

Figure 3: number of patients entered each year in the database. Different grey levels correspond to different age categories. See legend for explanation.



Diagnosis

Among the 602 patients and 1727 observations in the database, 743 different strings of characters were used to qualify the patients diagnosis. This number essentially reflects the absence of standardisation for this variable in the present database. This will have to be corrected in the forwarded database development to

allow for a more interesting use of this information in the future. In fact, this part of information, on which most studies rely for their database extraction is certainly not yet appropriately coded and future reflexion on an adequate and efficient coding is still in progress.

Karyotypes.

The karyotypes distribution according to the database variable *status* is given in table 4.

Table 4: Karyotypes distribution according to the database variable *status*.

Status code	10	15	16	25	35	45
Karyotype category		Normal		Abnormal		No mitose
Number of patients (%)	1 (0.2%)	233 (38.7%)	1 (0.2%)	151 (25.1%)	157 (26.1%)	59 (9.8%)

The karyotype variable takes 281 different values for the 1727 observations contained in the database. Among the 602 patients, 204 (33.9 %) have at least one observation different from NA, 46,XX and 46,XY. The age distribution for these patients is summarised in table 5.

Table 5: Summary of age distribution for patients with at least one karyotype problem.

Number of patients	204
Number of NA	1
Mean	63.3
Standard deviation	16.8
Minimum	0
Median	67
Maximum	93

Fifty-one patients have chromosome 5 involved in their karyotype description. The age and sex distribution of these patients is given in table 6.

Table 6: Summary demographics description for patients with involved chromosome 5:

Number of patients	51
Number of females	30
of males	21 (sex-ratio not significantly different from 1)
Age mean	65.18
Age standard deviation	15.9
Age minimum, median, maximum	8 68 85

Fifty-three patients have chromosome 7 involved in their karyotype description. The database of these patients is given in the appendix 4 with all databases. The age and sex distribution of these patients is given in table 7.

Table 7: Summary demographics description for patients with involved chromosome 7:

Number of patients	53
Number of females	22
of males	31 (sex-ratio not significantly different from 1)
Age mean	60.40
Age standard deviation	16.8
Age minimum, median, maximum	8 64 85

Appendix 1: Database codes

Variable inst

A:	Antwerpen	MG:	Mont-Godinne
B:	Brugge	N:	Namur
BP:		P:	Périphérie
C:	Charleroi	PL:	
G:	Gent	R:	Gasthuisberg
J:	Jolimont	RG:	
L:	St Luc	RH:	Gasthuisberg-Hemato
LS:		RP:	Gasthuisberg-Pediatrie
LX:	Luxembourg		
MB:			

Variable Tissue.1

500	Blood
501	Rate
505	Bone marrow

Variable Tissue.2

Sample number of the considered tissue as defined by tissue type.1. For example,
500/1: Blood, first sample
500/7: Blood, seventh sample

Variable status

15	Normal karyotype
16	Request for new sample
19	Not performed
25	Abnormal karyotype
35	No mitose
45	Normal with comment

Annex III

MDS: Analysis of historical database - extended confirmed database

Author:

Dr. Alix Collard

Investigators:

Dr Karel Van Damme

Dr. Ludwine Casteleyn

Dr. Alix Collard

Dr. Lucienne Michaux

Prof. Dr. Anne Hagemeyer

Prof Dr. Herman Van den Berghe

MDS: Analysis of historical database:

Analysis for the patients from the extended confirmed database
diagnosed in 1980 and after

Introduction

This report concerns 793 patients who have been referred to the Center for Human Genetics (Leuven) between 1980 and 1996, and who give rise to 2412 records. These records were selected from the root database extracted from the CME Oracle database according to the following inclusion/exclusion criteria:

- tissue type 500, 501 or 505
- diagnostic after 1/1/1980
- patients younger than 15 excluded
- diagnostic containing one of the following character sequences (where a * might represent any series of characters and ? represents any character):

MDS	*RAEB*
SMD	*R.A.E.B*
M.D.S	*CMM*
S.M.D	*R.A*
D?SPL	*REFRAC*
D?SMYELOP	*CYTOPENIE*
D?SERETHROP	*AREB*
RA	*LEU?OPENIE*
RAE	*THROMBOCYTOPENIE*
PANCYTOPENIE	*A.R.E.B*
C.M.M	*LMMC*
L.M.M.C	*ACHRES*
ACRES	*A.S.I.A*
A.I.S.A	*C/*

These different strings correspond to different possible qualifications of diagnostic compatible with MDS, except for «C/ » which is a qualificator given for patients with confirmed myelodysplasie.

- patients referred as MDS cases by the contact physicians but not originally included in the root database (see annex 1 and 2 for details) are added to this database

The selection was made in order to give more confidence that these selected patients diagnosis were indeed MDS. The selection criteria included the patients with a confirmed myelodysplasia (« C/ » in the diagnosis) plus patients confirmed by clinician. This selection gives little more weight to hospitals who have sent confirmation lists(namely Mont-Godinne, St Luc and Jolimont), but this is not expected to bias the population demographic or genetic structure.

The information collected on these patients and relevant for this study is described in 16 variables which are summarised in table 1.

Table 1: Variable list for the MDS retrospective study database.

Variable name	variable content
Patid	patient identification number
Sex	Sex
Zip	zip code (residence)
Age	Age
Birth.date	birth date
Sample.number	sample number
Sample.date	sample date
Inst*	Hospital
Tissue.1*	tissue type. first sample
Tissue.2*	tissue type. second sample
Diag	Diagnostic
Lab	lab number
Clone	clone number
Number.cell	number of cells within one clone
Karyo	Karyotype
Status*	status code

*See appendix 1 for codes

Data quality

Concerning the demographic variables, some errors were detected especially concerning the variable age as given by the database variable. Indeed, some patients were specified as having a negative age. This is due to the incomplete coding of the birth year, which raises problems for people who were born before 1900. Corrections have been made in the database for all patients with negative ages. Moreover, all checks have been made to check birth dates for all patients with coded ages less than 20 years. This didn't induce any further correction. The list of all patients with age less than or equal to 20 years (according to the initial coding) is given in appendix 3.

An anomaly has also been detected for patient 87 782 for whom one observation is dated from 1977, but not classified following the chronological order as usually in the database. This patient observations were sorted by sample date to correct for this irregularity.

Results

Demographics

The whole population considered in this report comprises 466 males (58.8%) for 324 females (40.8%), 3 patients with no information available for sex. The sex ratio is significantly different from 1 (p -value < 0.0001). This finding is similar to the result of Dr. C. Nisse who found a preponderance of male patients (62%) (Nisse, 1996), as is usually admitted (Aul et al, 1995).

The age distribution for the study population is summarised in table 1 and a histogram is given in figure 1. Ninety-seven patients (97, 12.2%) are less than 50 years old at inclusion in the database, and 304 (38.1%) are less than 65 years old. This age distribution seems to conform with those usually reported. For example, Dr. C. Nisse (Nisse, 1996) reported a mean age of 67.3 ± 12.5 years (p -value for two-sided comparison of both sample age characteristics: 0.47). Another study (Third MIC Co-operative study group, 1987) reported a median age of approximately 70 years and another reported a median age of 73 years with only 11 patients among a cohort of 141 under 60 years of age (Mufti et al, 1985). It is usually admitted that more than 75% of the patients are older than 60

years at diagnosis (Aul et al, 1995). In our own population, 583 (73.5%) patients are more than 60 years old.

Table 1: Summary of age distribution for population database.

Number of patients	793
Number of NA	5
Mean	66.47
Standard deviation	14.36
Minimum	15
Median	69
Maximum	100

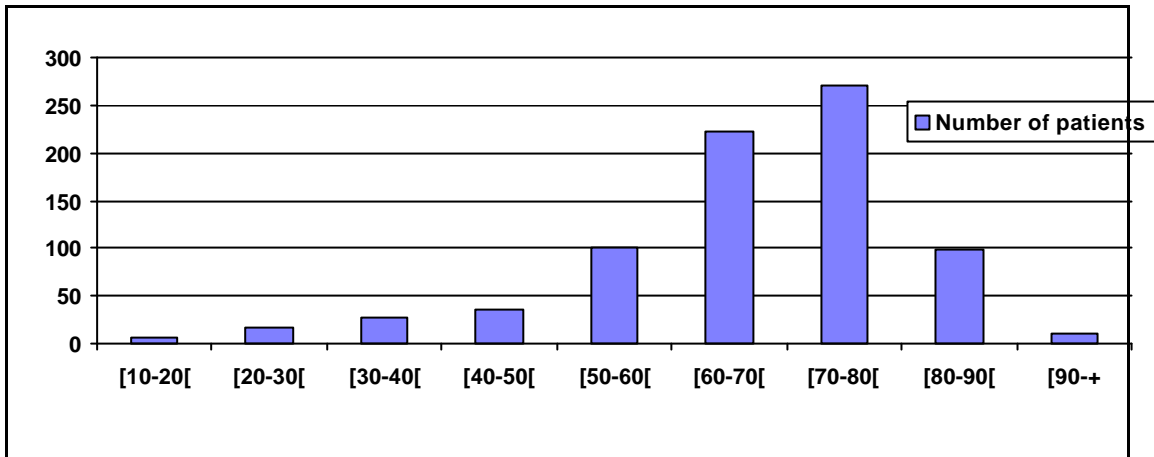


Figure 1: Age distribution

Hospital distribution.

The number of patients included by hospital is given in table 2. Given the database construction strategy, this is only for descriptive purpose, since of course, St Luc and Gasthuisberg are over-represented.

Table 2: The first line gives the institution abbreviation as used in the database, while the second line gives the corresponding patients entered in the database by this institution.

Inst.	NA	A	B	G	J	L	LS	LX	MB	MG	N	P	R	RG	RH
Nb.	9	8	19	13	67	275	4	2	4	102	2	195	8	2	83

Time evolution

The number of patients newly entered each year is described in figure 2. Time evolution is crossed with hospital to provide table 3.

Figure 1: number of patients entered each year in the database

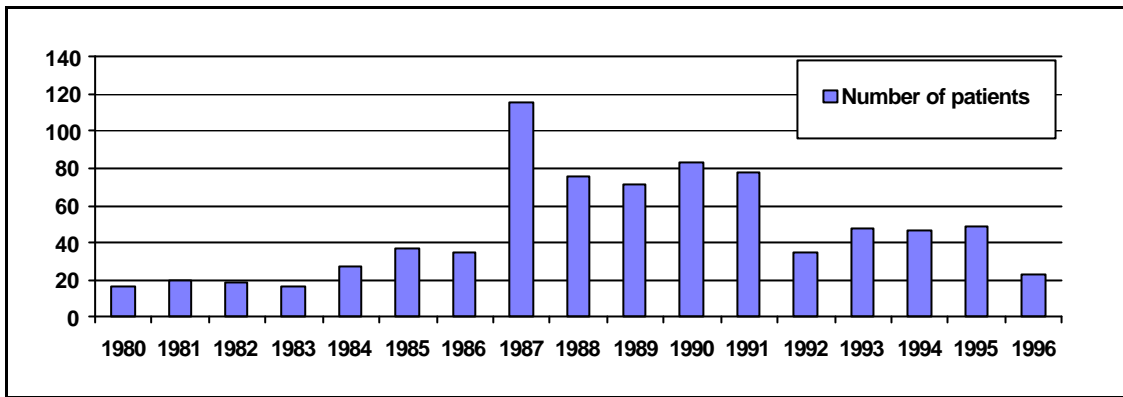


Table 3: Hospital by-year number of patients.

	A	B	G	J	L	LS	LX	MB	MG	N	P	R	RG	RH
1980	0	0	0	0	9	0	0	2	5	0	0	0	0	0
1981	0	0	0	0	12	0	0	1	5	1	0	0	0	0
1982	0	0	0	0	11	1	0	1	5	0	1	0	0	0
1983	0	0	0	0	14	1	0	0	1	0	0	0	0	0
1984	0	0	0	0	24	0	0	0	3	0	0	0	0	1
1985	0	1	1	0	27	0	0	0	3	0	0	0	0	5
1986	0	2	0	0	22	0	0	0	4	0	4	0	0	3
1987	6	11	8	0	36	0	0	0	15	0	35	0	0	3
1988	0	2	4	0	22	1	0	0	10	0	33	0	2	1
1989	1	1	0	5	17	1	0	0	10	1	14	0	0	20
1990	0	1	0	16	10	0	2	0	8	0	25	0	0	19
1991	1	0	0	16	16	0	0	0	8	0	15	4	0	17
1992	0	0	0	5	13	0	0	0	5	0	8	0	0	3
1993	0	0	0	6	17	0	0	0	8	0	10	1	0	5
1994	0	0	0	12	11	0	0	0	2	0	15	2	0	4
1995	0	1	0	4	11	0	0	0	5	0	24	1	0	2
1996	0	0	0	3	3	0	0	0	6	0	11	0	0	0

Karyotype analysis:

The database for karyotypic analysis is extracted from database with 2412 records on 793 patients. Patients with only blood sample leading to «normal» status were excluded from this database. Moreover, 3 patients present a constitutional anomaly T(13;14) which is not taken into account. This leads to a database of 2361 records for 750 patients. Among these 750 patients,

- 97 have only blank or NA karyotypes,
- 381 have a «normal karyotype» (no anomaly observed and at least one observation which is either 46,XX or 46,XY)
- 272 have at least one anomaly (at least one anomaly observed)

The age distribution and sex repartition for the 750 patients may be described as:

Age:	mean± sd:	66.22 ± 14.50
	min-med-max:	15-69-100
Sex:	305 F, 442 M, 3 unknown	

The comparison of sex repartition across patients with and without anomaly gives the following table:

Sex	with anomaly	without anomaly	P-value*
Female	107 (39.3%)	157 (41.2%)	0.6899
Male	165 (60.7%)	224 (58.8%)	

GENERAL KARYOTYPE ANALYSIS

If we divide the sample of patients with available karyotypic information in the following three groups: the patients with a normal karyotype, the patients with a simple abnormal karyotype (less than 3 chromosomal anomalies) and the patients with a complex karyotype (3 or more karyotypic anomalies), the following table might be built.

Age characteristics	Complex karyotype	Simple karyotype	Norma karyotype
Sex F/M	33 / 40	74 / 125	157 / 224
number of patients (number of NA)	73 (1)	199 (0)	381 (2)
mean \pm sd	63.79 \pm 15.56	64.45 \pm 16.28	67.30 \pm 13.28
min-med-max	16 - 66.5 - 85	15 - 67 - 100	15 - 70 - 93
p-value*	0.0878	0.0847	

* Wilcoxon rank-sum test for the comparison towards the patients with a normal karyotype

Among the patients with a complex karyotype, there are 21 patients whose karyotype does not involve chromosome 5 nor 7.

KARYOTYPE ANALYSIS PER CHROMOSOME:

CHROMOSOME 5

73 patients with at least an anomaly involving chromosome 5, among whom 43 F and 30 M with the following age distribution summary characteristics [73 (1); 65.19 \pm 15.24; 18-68-85] corresponding to [number of patients (number of patients without information); mean \pm sdev; min - med - max]

For the following analyses, the first anomaly involving the chromosome of interest has been taken into account.

26 patients have a « 5q- » as sole anomaly, while 21 have this 5q- combined with other anomalies, and 18 have a « -5 » for anomaly of the chromosome 5. The remaining 8 patients have different other anomalies involving chromosome 5. Sex and age distribution characteristics are summarised in the following table:

	5q- alone	5q- plus other	-5	other 5
Sex: F / M	19 / 7	11 / 10	9 / 9	4 / 4
Age Nb (Nb of NA)	26 (1)	21 (0)	18 (0)	8 (0)
mean \pm sd	64.60 \pm 17.45	66.05 \pm 11.71	66.72 \pm 14.01	61.37 \pm 20.46
range	20 - 67 - 85	40 - 65 - 83	24 - 65 - 84	18 - 71 - 79

CHROMOSOME 7

69 patients with at least an anomaly involving chromosome 7, among whom 32 F and 37 M with the following age distribution summary characteristics [69 (0); 60.13 ± 17.95; 16-65-85]

For the following analyses, the first anomaly involving the chromosome of interest has been taken into account.

11 patients have a « 7q » as chromosome 7 anomaly, while 48 have a « -7 » for anomaly. The remaining 10 patients have different other anomalies involving chromosome 7. Sex and age distribution characteristics are summarised in the following table:

	7q-	-7	other
Sex: F / M	1 / 10	24 / 24	7 / 3
Age Nb (Nb of NA)	11 (0)	48 (0)	10 (0)
mean ± sd	66.36 ± 13.37	56.79 ± 19.28	69.30 ± 9.65
min-med-max	35 - 65 - 85	16- 60 - 84	52 - 68.5 - 83

CHROMOSOME 8

67 patients with at least an anomaly involving chromosome 8, among whom 22 F and 45 M with the following age distribution summary characteristics [67 (0) ; 65.78 ± 14.99 ; 18 - 68 - 87]

For the following analyses, the first anomaly involving the chromosome of interest has been taken into account.

57 patients have a « +8 » as chromosome 8 anomaly, while 10 have different other anomalies involving chromosome 8. Sex and age distribution characteristics are summarised in the following table:

	+ 8	other
Sex: F / M	19 / 38	3 / 7
Age Nb (Nb of NA)	57 (0)	10 (0)
mean ± sd	66.5 ± 15.63	64.20 ± 11.21
min-med-max	18 - 69 - 87	44 - 67 - 80

CHROMOSOME 20

44 patients with at least an anomaly involving chromosome 20, among whom 15 F and 29 M with the following age distribution summary characteristics [44 (1); 64.74 ± 13.72; 31-83]

For the following analyses, the first anomaly involving the chromosome of interest has been taken into account.

29 patients have a « 20q- » as anomaly involving the chromosome 20, while 5 have a « -20 » for anomaly. The remaining 10 patients have different other anomalies involving chromosome 20. Sex and age distribution characteristics are summarised in the following table:

	20q-	-20	other
Sex: F / M	10 / 19	2 / 3	3 / 7
Age Nb (Nb of NA)	29 (0)	5 (0)	10 (1)
mean ± sd	63.41 ± 13.85	66.20 ± 9.50	68.22 ± 15.82
range	31 - 83	54 - 76	43 - 83

CHROMOSOME 11

Nineteen patients have a simple anomaly involving chromosome 11 (without involving chromosome 5 nor 7) , among whom 7 females and 12 males and with the following age distribution summary characteristics [19 (0) ; 67.84 ± 13.34 ; 33 - 68 - 87]

Annex IV

MDS case-control study: Study Protocol

Contents:

1. Introduction
 2. Objectives
 3. Investigator and study site
 4. Informed consent, confidentiality and ethics
 5. Materials and methods
 6. Statistical considerations
- Bibliography
Appendix 1-4

Investigators / Authors:

Dr Karel Van Damme

Dr. Ludwine Casteleyn

Dr. Alix Collard

1. Introduction

Myelodysplasia (MDS) is a clonal haematopoietic disorder, arising as a consequence of haematopoietic stem cell damage. From the existing literature (Third MIC Co-operative group, 1988, Aul et al., 1995, Rodella et al., 1993), the hypothesis was raised that in a number of MDS cases, a genome attack might be linked with a specific exposure of the patient. Case-control studies or case reports have incriminated antineoplastic drugs (mainly alkylating agents) (Pedersen-Bjergaard et al, 1990, Levine and Blomfield, 1992), ionising radiation's (Mole, 1990), benzene exposure (Aksov, 1985, Brandt, 1992) and other solvents (Brandt, 1992), pesticides and herbicides (Brown et al, 1990, Richardson et al , 1992), ethylene oxide (Hogstedt, Aringer and Gustavsson, 1992), chloramphenicol (Schmitt-Graff, 1981), radon (Henshaw, Estough, Richardson, 1990) and electromagnetic fields (Henshaw, Estough, Richardson, 1990, Coleman, Beral, 1988). However, apart from alkylating agents, ionising irradiation and benzene, a role for occupational or environmental exposure has not been demonstrated in the aetiology of MDS until recently (Cartwright, in Mufti & Galton eds, 1992). Five recent case-control studies addressed this issue. Farrow et al. (Leukemia, 1989) found a higher incidence of exposure to ammonia, petrol and diesel fumes in MDS than in controls, based on 60 patients. Goldberg et al. (Cancer Res, 1990) found more frequent exposure to pesticides in MDS than in controls, based on about 50 patients. West et al. (Leukemia research, 1995) performed a case-control study on 400 individually matched pairs and showed associations between radiation and myelodysplasia. Nisse et al. (1995) performed a case-control study on 204 patients and 204 general population based controls. A significant excess risk of MDS was observed in in male machine operators, agricultural workers, textile workers and textile machine operators, sales representatives, health professionals, and persons with gardening as a hobby. They also demonstrated an excess risk for exposure to stone dust, cereal dust, cotton and flax dust, ammoniac, oil, metal fumes and dust, solvents, fertilisers, poultry, livestock and to septic risk factors. Many of these findings are suggestive of a correlation with exposure to pesticides. Also living near an industrial plant, tobacco consumption and smoking were significant risk factors, while exposure to diagnostic X-rays was not. Ido et al. (1996) conducted a case-control with 106 MDS patients in Japan and showed increased risk for exposure to alcohol, and increased but not significant risk for exposure to cigarette, hair dye products and organic solvents.

A correlation between exposure to known mutagenic and carcinogenic agents and chromosomal aberrations in the neoplastic cells of patients with haematological malignancies was shown in several studies (Mitelman et al, 81; Golomb, 82; Brandt, 83; etc) but could not be demonstrated in others.

Inherited differences in metabolic capacity (polymorphism's in drug-metabolising or detoxifying enzymes), DNA repair capacity, and genetic predisposition to specific diseases are accounting for (part of) the wide variation in individual responses to environmental agents. Evidence exists on excess of GSTT1-null genotypes among myelodysplastic syndrome patients (Chen et al, Lancet) and a higher frequency of NQO1 inactivating polymorphism's has been identified in subjects developing haematotoxicity after benzene exposure.

2. Objectives

This case-control study intends both to test hypotheses and to generate hypotheses. Pesticides exposure will be particularly emphasised in this report, combining information from every possible source of exposure, either professional, environmental or other.... In addition, exposure linked with drug and medical imagery are particularly detailed. Attention will also be paid to the study of the latency period between any particular exposure and the MDS diagnosis.

Unlike the usual epidemiological studies such as the ones just evoked in the introduction, the emphasis in this study will be put on the reliability and quality of the collected information, rather than on the standardised way of collecting information and the quantity of subjects. The rationale for this different option resides in the belief that characterising an individual exposure profile using any form of standardised questionnaire is likely to result in a partly incorrect characterisation due for instance to the general population's vague knowledge about exposures. A different approach will thus be developed for the exposure history characterisation, the main objective being to be able to assess the "real" exposure as precisely as possible through a common "reconstruction" of the patients exposure history by a structured conversation between interviewee and interviewer, and thus not by an endless series of questions. In this approach, the interviewer must have had medical training and be very well trained in occupational and environmental health, because he or she will have to make an exposure assessment based upon what will be said during the interview. This strategy is expected to lead to a more precise and complete exposure characterisation for both cases and controls, thus resulting in more reliable data for comparison of the exposure experience of cases and controls. The final idea is that, unlike in studies with standardised questionnaires, where real effects might not be detected (except in situations where the effect is big enough) due to some mixing up of correct and incorrect exposure data which artificially increases the observations' variance, the higher level of precision reached through more appropriate interviewing techniques should allow to reduce this variance and even to reach conclusions based upon smaller sample sizes.

Developing hypotheses concerning the latency periods between exposure and MDS diagnosis is another challenging issue for this study. In order to meet this goal, any exposure will be detailed with respect to its time location, according to the decade before diagnosis when it occurred. Hence period 1 is defined as the decade preceding diagnosis, period 2 corresponds to the second decade before diagnosis, and similarly periods 3 and 4 to the third and fourth decades before diagnosis.

Besides, particular scales will be developed in order to be able to divide subjects into categories of « pesticide exposure », « drug exposure », « solvent exposure ».

Finally, the exposure level within patients presenting a cytogenetic anomaly involving chromosome 5, 7 or 11 were computed, as these are the cytogenetic alterations most commonly found in MDS patients.

3. Investigator and Study Site.

The study is conducted and coordinated by Dr. Ludwine Casteleyn and Dr. Karel Van Damme. Both are affiliated to the Center for Human Genetics Gasthuisberg, University Hospital of the University of Leuven and to the University of Antwerp, Department of Epidemiology and Community Medicine. Statistical analyses will be performed by Dr Alix Collard, working at the FUNDP. The final results report will be written in collaboration between all the participating investigators.

4. Informed consent, confidentiality and ethics.

This study will be carried out in accordance with the declaration of Helsinki.

The investigator will obtain individual's consent before entering him or her into the study using the form presented in appendix 1. In accordance with all regulations concerning privacy protection, the information collected from each individual will be held in confidence: The participant's name will be removed from documents and replaced by a code, and after the data are transferred to computer files, the original documents will be stored in locked locals, with access to files and storage made available for authorised staff only. The data will be kept as long as they are useful for further research.

Approval by the Ethics Committee of the hospitals of St Luc and Gasthuisberg, which are directly involved in this study was obtained for this study as well as for the informed consent form.

All MDS cases and control patients are contacted in a way which maximises the involvement of treating physicians, both in the hospital and the patient's general practitioner.

5. Materials and methods.

5.1. Study design.

Given the disease rarity, a case-control was of course the design of choice. Matching on age and sex was considered important since these factors are known to be associated with both disease (the sex ratio for MDS is usually accepted to be 1.2, and myelodysplastic syndrome is known to occur much more frequently in older persons) and suspected exposure (professional exposures are known to be different according to the sex and age of a person). Additionally, matching for these factors should allow for a similar potential of recall of past exposure for cases and controls. Moreover, as we are also interested in studying the relationship between chromosomal aberrations or genetic modifications otherwise, matching for gender is absolutely necessary not only for apparent theoretical reasons, but also because we have learnt from former genetic analyses of cases that there is an unequal distribution of types of chromosomal aberrations between males and females. A large discussion about a possible matching on social status took place, but the final decision was not to match with respect to this characteristic, because the use of hospital-based controls is likely to lead to the same catchment population for both cases and controls on the one hand, and because social status is a very difficult concept which might be formalised in different ways and for which an approach which is both practical and reliable is likely to be very difficult to achieve.

Cases known to be associated with a particular hereditary susceptibility to cancer and who may bias the study results will be excluded. Patients who are unable to participate during the interview period will be excluded. The interviewing approach considered is not compatible with the theoretical concept of 'blinding'.

5.1.1. Study power considerations

A sample size of 30 case-control pairs is forwarded for the study, given the very time-and-person-consuming interviewing and data management process.

Given the context of 1:1 matching design, sample size computations based upon the recommendations of J.M. Lachin (1992) in his paper about " power and sample size evaluations for the Mac-Nemar test with applications to matched case-control studies " lead to a number of 20 discordant pairs to be able to detect an odds-ratio of 4 with 80% power. The minimal detectable effect is thus likely to be an odds ratio of 4, with an odds ratio of 3 being detectable with 29 discordant pairs (Table 1).

Table 1: number of discordant pairs to be observed to provide 80% power to detect an odds ratio ω at the usual 0.05 level is given in the following table for different possible values of ω :

w	1.5	2	3	3.5	4
Number of discordant pairs	194	69	29	23	20

5.1.2. Study duration

Patients diagnosed between 1/1/1996 and 31/12/1998 will be considered for this study, and the interviews of the cases and their controls are planned to take place in 1998. This could imply that some of the cases and controls have died before the interviews are taken. For practical reasons, this was the only possible time schedule. There is no major reason to believe that this might bias the results.

5.2. Study population.

5.2.1. Disease definition:

Our interest focuses on primary MDS , i.e. MDS secondary to anti-cancer treatment will be discarded. All patients diagnosed between 1/1/1996 and 31/12/1998 in one of the hospitals participating into the study as MDS cases according to the French-American-British (FAB) diagnostic criteria (Bennet et al., 1982), and aged no less than 20 and no older than 74 years at diagnosis will be included, provided he/she accepts to give his/her informed consent. MDS specification will be based upon the FAB criteria.

5.2.2. Exposure definition:

In relation with the main objective of this study, and in view of reported and planned research on the same subject by other research teams, the following exposures will carefully be assessed.

- residence and work environment
- occupational history
- drug use
- medical imagery experience
- tobacco consumption
- drinking habits
- pesticide exposure
- solvent exposure
- dioxin exposure

by occupational and environmental health medical experts based upon the answers collected during the interviews. Details about the scales used are described in a following section.

5.2.3. Case definition and selection.

In order to be validly included as a case, a patient will have to conform the following requirements:

- MDS case diagnosed between 1/1/1996 and 31/12/1998 (the rationale behind this requirement is that recent diagnosis will be more uniform, recall is expected to be more accurate and is more certain to concern exposure prior to diagnosis, and decease due to the disease of concern is limited;

- diagnosis between the age of 20 or more and less than 75 (the rationale between the age limits is that patients aged no younger than 20 and younger than 75 are likely to have experienced at least some professional exposure and are hoped to experience less recall problems than older patients, while patients aged 75 or older are likely to experience recall problems, and patients less than 20 won't have worked);
- diagnosed in accordance with the FAB diagnostic criteria;
- cases secondary to anti-neoplastic agents and patients unable to participate during the interview period will be excluded (the causative factor of these MDS's is thought to be related to the anti-neoplastic treatment already received).

5.2.4. Control definition and selection

Different possibilities for control definition were considered. For each of these possibilities, the advantages and the drawbacks were identified, with special reference to the risk for overmatching for the exposures of concern. Ideally, several controls each responding to a different definition should be used. This may allow to study more systematically the correlations found. For practical reasons, this idea was abandoned.

Given that some of the exposure factors under study are thought to be linked with the environment, it was felt that the control population should be chosen in a geographical zone big enough to allow for a possibly different environmental exposure. In other words, controls in the same street or village –for instance having the same general practitioner- were not considered appropriate because it was felt that for instance , farms as living environment tends to cluster within villages which would require a very sophisticated and hardly feasible assessment for pesticide exposure. Or in more industrial areas, people may have experienced the same occupational exposures.

Sampling controls from electoral lists was considered inappropriate for several reasons, the most important reason being that it appeared impossible to know the catchment population of the hematology services participating in the study. Subsequently it was impossible to ponderate the selection of controls as a function of the geographical distribution of the catchment population.

Hospital-based controls were thought to offer a valuable alternative, even if the catchment population may be different for different diseases. This problem is assumed –although not proven- to be overcome by including different pathologies as potential controls. Another disadvantage is that other pathologies may be related to similar exposure factors, and these correlations are often unknown. An advantage of hospital based controls may be that the interests in recalling exposure situations is similar to that of the cases.

A case-case study with mixed pathologies was the final option.

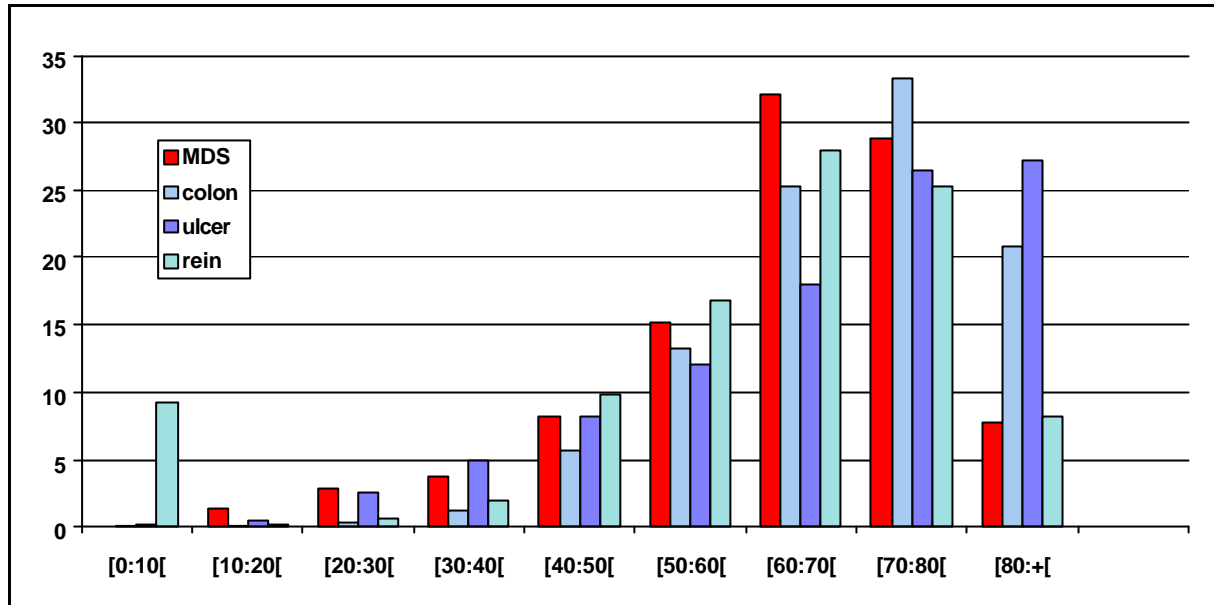
Controls will be selected from the pooled population of

- renal cancer
- colon cancer
- gastro-duodenic ulcer

diagnosed between 01/01/1996 and 31/12/1998 in the same centre as the case. Diagnostic criteria for these diseases are described in appendix 2. These three diseases have been selected since the patients they affect share approximately the same age structure as the MDS patients as is illustrated in figure 1, bases upon data of the public health administration of the Netherlands. The use of multiple diagnoses protects against the possible failure to detect an effect of exposure should the MDS and a single control-disease be both associated with this effect, on the one hand, and also for the claiming that an exposure affects the risk of MDS, when in fact the effect of exposure is really being linked to the single diagnosis from which the controls are drawn, on the other hand.

Figure 1: Age structure of MDS cases as determined from the retrospective study performed on the sample of patients with confirmed diagnosis and referred to the CME for cytogenetic analyses between 1980 and 1996, and age structure of renal cancers, colon cancers and gastric ulcers populations, as based upon data of the public health administration of the

Netherlands.



For every MDS case and for each of the three reference pathologies, the last 2 persons who were either hospitalised or seen at the consultation will be selected from the medical archives, amongst the cases who are in compliance with the above criteria. These selected cases will be ranked in arbitrary order by the blinded researchers and subsequently contacted in that order until someone agrees to be interviewed. As soon as the total number of 10 interviewed referent persons with the same pathology is reached, the researchers will first select their referent amongst the other pathologies.

5.3. Rating instruments.

A face-to-face interview will be conducted by either Dr. Karel Van Damme or Dr. Ludwine Casteleyn (or both) at the hospital or at the patient's home according to the patient's best convenience using an apparently informal but well structured conversational approach which intends to allow a common reconstruction of the patient's exposure history so that the measures of interest will be determined as accurately as possible and systematically registered on a registration document, on which the exposure history for the factors of interest is reconstructed on a time scale.

The variables which are to be measured are described in appendix 4 and include demographic data, residence during the 40 years preceding the diagnosis, medical history together with drugs, medical imagery, operations and past illnesses, smoking habits, diet, occupation and hobbies, solvent exposure, pesticide exposure, dioxin exposure, ... The period concerned by the information collected will thus also be recorded, according to the decade preceding the diagnosis to which it belongs. This information will be compared across cases and controls for socio-economic characteristics, demographic data, environmental exposure (housing environment, water source, heating medium, humidity problem, renovation work, neighbourhood), medical history (diseases, medical imagery, drugs, toxicology, chemotherapy), drinking habits, tobacco consumption, occupational exposure, For pesticides exposure, solvent exposure and dioxin exposure, a strategy for exposure assessment will be followed which starts from the data collected. A series of questions have to be answered in different parts of the questionnaire. These data will be brought together for each interviewee. Exposure situations and histories of the interviewees will be mutually compared and ranked independently of status. This is an ad hoc developed method which will be tried out in this study and is expected to minimize the loss of information when ranking cases and controls with respect to their exposure histories. For the analysis, pooling in categories will be done.

5.3.1 Mode of interview

MDS cases and control (referent) patients will be interviewed by physicians who are epidemiologists specialised in occupational and environmental health. The interviewing strategy will consist of semi-open questioning in which the exposure history of the last 40 years previous to the diagnosis will be reconstructed in common with the interviewee and result in the interviewer filling in the questionnaire presented in separate annexes V and VI (not in appendix). This document can be considered as a questionnaire which the interviewers have to provide with the best possible answer to each question. This interviewing strategy is expected to provide more accurate

answers than the more classical strategy where the direct answers provided to a list of questions are simply recorded. In addition, it is unclear how such systematic reconstruction could be done through strictly standardized questioning. The questionnaire is divided into a general questionnaire and a series of specific questionnaires to be used for specific occupational activities. The interview will be conducted such that it may be perceived by the patient as a conversation, and not as an endless series of question.

5.3.2 Scales

5.3.2.a. Available scales

The living environment is described using the following scale:

- 1:city (a=open space, b=border, c=centre)
- 2:village (a=open space, c=centre)
- 3:country site
- 4:forest

The following scale is also used for keeping the information about a possible tropical environment during one of the decades:

- 0: non
- 1: occasionally
- 2: living

Education is described using the following two codes:

General education		Supplementary education	
Primary school:	code 1	Yes:	code 1
Secondary school:	code 2	No:	code 0
Professional/Technical formation:	code 3		
Superior short-type:	code 4		
Superior long-type:	code 5		

Coffee consumption is expressed in cups of light coffee using the following correspondence:

One cup of strong coffee is considered to correspond to three cups of light coffee, and one cup of normal coffee is considered to correspond to 2 cups of light coffee. One liter of coffee is considered to correspond to nine cups of coffee.

Tobacco consumption is described using the following scale:

- 0:never
- 1:passive(occasion.)
- 2:passive(considerable)
- 3:occasional
- 4:in company or after meal
- 5:with filter
- < 1/2 pqt /day
- 6:without filter
- < 1/2 pqt/day
- 7:with filter
- [1/2-1]pqt/day
- 8:without filter
- [1/2-1]pqt/day
- 9: > 1 pqt/day
- 10: > 2 pqt/day

Occupations will be described using ILO ISCO-68 codes and NACE-codes.

5.3.2.b. Scales elaboration

For the pesticide exposure characterisation, as well as for the solvent exposure and for the dioxin exposure, the choice is to elaborate the scale after the data collection, based upon the exposition profiles observed in the study, just by an ordering of all profiles which will allow to attribute a score to each patient.

6. Statistical considerations.

6.1. Data management.

6.1.1. Database elaboration

Whereas the questionnaire is designed to be useful during the process of information collection, the database is designed to translate this collected information in an efficient way into a data structure which can be exploited for statistical analysis. Hence, all the information collected during the interview is translated into a series of variables, either of character type, or using numerical formats or date formats. The list of variables and associated codes and formats is presented in appendix 3.

6.1.2. From the questionnaire to the database

For most of the information collected, such as birth date, sex, ... the information collected during the interview can be translated directly into the corresponding variables. However, some coding instructions have been used in some specific situations which are more thoroughly detailed in appendix 4.

6.1.3 Database report

The main objective of this report is to provide some information about the database quality, giving information about the database contents (number of patients included, number of variables per patient), about possibly erroneous, illegal, or missing data. It also describes the patients included/excluded from statistical analyses, with the reasons for the exclusion.

6.2. Statistical analysis.

Usual summary statistics will be provided for variables description, Wilcoxon test will be used for continuous variables comparison, while Fisher's exact test will be used for comparison of binary variables. Matched pair analysis will be performed using the Mc Nemar test for binary exposure comparisons and odds ratio will be computed.

For each exposure, the analysis will proceed from the very detailed to the more grouped level, thus analysing separately the four decades, before coming to the exploitation of the whole study-period exposure, and exploiting each variable separately before coming to some possible combinations of variables.

In the final statistical report, the number of pairs of MDS patients with their matched control will be provided, together with detailed information about the control pathologies. The number of patients having refused the interview will also be given, and the possibility of bias introduction assessed.

Median age and sex repartition will be described for cases and controls. In addition, the repartition of the cases in the different FAB categories will be detailed. Final study level will be presented and compared between cases and controls.

Demographic characteristics (number of child, height, weight, body mass index) will also be described for cases and controls and compared.

Environmental exposure will be described and compared for the housing environment, the source of the water used, the heating media, the exposition to humidity, to renovation work, and to neighbourhood possible nuisance.

Medical history will also be described and compared, with more attention brought to drugs consumption and medical imagery use.

Exposure linked with the alimentation and drinking will also be studied, by a description and comparison of coffee, tea, caffeinated and non caffeinated sweet drinks, milk, wine, beer and alcohol consumption.

Histories of tobacco consumption will also be studied in detail.

Occupational exposure will be analysed by job titles. The mean number of jobs held for at least 6 months will be computed and compared.

Pesticide, solvent and dioxin exposure will be detailed and compared using the index derived from the different profiles orderings.

Bibliography

- [1] Aksov M (1985) Malignancies due to occupational exposure to benzene, *Am F Ind Med*, 2:217-245.
- [2] Aul C, Gattermann N, Schneider W (1994) Epidemiological and etiological aspects of MDS, *Leukemia and Lymphoma*, 16:247-262.
- [3] Bennet JM, Catovski D, Daniel MT et al (1982) Proposals for the classification of the myelodysplastic syndromes, *British Journal of Haematology*, 51:189-199.
- [4] Brandt L (1992) Exposure to organic solvents and risk of haematological malignancies, *Leukemia Res*, 16:67-70.
- [5] Brown LM et al (1990) Pesticide exposures and other risk factors for leukemia among men in Iowa and Minnesota, *Cancer Res*, 50:6585-6591.
- [6] Cartwright RA (1992) Incidence and epidemiology of the myelodysplastic syndromes, in: Mufti G, Galton D (eds). *The Myelodysplastic syndromes*. Churchill Livingstone, London, 23-32.
- [7] Coleman M, Beral V (1988) A review of epidemiological studies of the health effects of living near or working with electricity generation and transmission equipment, *Int J Epidemiol*, 17:1-13.
- [8] Farrow A, Jacobs A, West RR (1989) Myelodysplasia, chemical exposure, and other environmental factors, *Leukemia*, 3:33-35
- [9] Goldberg H et al (1990) Survey of exposure to genotoxic agents in primary myelodysplastic syndrome: correlation with chromosome patterns and data on patients without hematological disease, *Cancer research*, 50:6876-6881.
- [10] Henshaw DL, Estough JP, Ricardson RB (1990) Radon as a causative factor in induction of myeloid leukemias and other cancer, *Lancet*, 1:1008-1112.
- [11] Hogstedt C, Aringer L, Gustavsson A (1986) Epidemiologic support for ethylene oxide as a cancer-causing agent, *JAMA*, 255:1575-1578.
- [12] Ido M et al (1996) A case-control study of myelodysplastic syndromes among Japanese men and women, *Leukemia research*, 20:9:727-731.
- [13] Levine EG and Bloomfield CD (1992) Leukemias and myelodysplastic syndromes secondary to drug, radiation and environmental exposure, *Semin Oncol*, 19: 47-84.
- [14] Mecucci C and Van den Berghe H (1992) Myelodysplastic syndromes, *Cytogenetics*, 6:3:523-541
- [15] Mufti GJ et al (1985) Myelodysplastic syndromes: a scoring system with prognostic significance, *British J of Haematol* 59:425-433.
- [16] Nisse C et al (1995) Exposure to occupational and environmental factors in myelodysplastic syndromes. preliminary results of a case-control study, *Leukemia*, 9:693-699.
- [17] Nisse C (1996) Syndromes Myelodysplasiques et exposition à des facteurs de risque professionnels ou environnementaux: Etude cas-témoin, dossier de fin de recherche, CERESTE
- [18] Pedersen-Bjergaard J et al (1990) Chromosome aberrations and prognostic factors in therapy-related myelodysplasia and acute nonlymphocytic leukemia, *Blood*, 76:1083-1091.
- [19] Richardson S et al (1992) Occupational risk factors for acute leukemia: a case-control study, *Int J Epidemiol*, 21:1063-1072
- [20] Rodella S et al (1993) Cytogenetics and occupational exposure in acute non-lymphocytic leukemia and MDS, *Scand J Work Environ Health*, 19:369-374.
- [21] Schmitt-Graff A (1981) Chloramphenicol-induced aplastic anemia terminating with acute nonlymphocytic leukemia, *Acta Haematol*, 66:267-273.
- [22] Third MIC Cooperative group 1987 (1988) Morphologic, Immunologic and Cytogenetic (MIC) working classification of the primary MDS and therapy-related MDS and leukemia, *Cancer Genet Cytogenet*, 32:1-10.
- [23] Van den Berghe H and Michaux L (1997) 5q-, twenty-five years later: a synopsis, *Cancer Genet Cytogenet*, 94:1-7.
- [24] West RR et al (1995) Occupational and environmental exposures and myelodysplasia: a case-control study, *Leukemia research*, 19:2:127-139.

Appendix 1: Draft consent form for interview (cases)

Code

TOESTEMMING

Naam

Adres

Hierbij stem ik er in toe om, in het kader van de studie met referentie **"ST/06/023"**, waarover ik alle gevraagde uitleg gekregen heb, ondervraagd te worden over mijn beroep, mijn gezondheid, mijn leefomgeving, en over de gezondheid van mijn familie in het algemeen.

Alle gegevens die ik verstrek zullen als confidentieel worden beschouwd en behandeld worden in overeenstemming met de Europese en/of de nationale wetgeving van de bescherming van persoonlijke gegevens.

Handtekening:

Datum

Handtekening Interviewer(s)

Datum

Dr Karel Van Damme
Dr Ludwine Casteleyn
Centrum voor Menselijke Erfelijkheid
Herestraat 49
3000 Leuven
Tel 016 34 62 20
Fax 016 34 60 63

Code:

CONSENTEMENT

Nom:
Adresse:

Dans le cadre de l'étude MDS, reference ST/06/023, sur laquelle j'ai reçu toutes les explications souhaitées, je consens par la présente à être interviewé/e au sujet de ma vie professionnelle, de ma santé, de mon environnement de vie ainsi que sur la santé de ma famille en général.

Toutes les réponses que je donnerai dans le cadre de cette étude resteront confidentielles et seront traitées conformément à la législation européenne et/ou nationale relative à la protection des données à caractère personnel.

Signature de l'interviewé/e

Date

Signature de l'interviewer(s)

Date

Dr K. Van Damme
Dr L. Casteleyn
Centrum voor Menselijke Erfelijkheid
Herestraat 49, 3000 Leuven
Tel: 016 34 62 20
Fax: 016 34 60 63

Appendix 2: Diagnostic criteria

The following ICD codes have been used for the controls selection in the hospital databases.

- 531.** Gastric ulcer
- 532.** Duodenal ulcer
- 189.0 Kidney neoplasm (primary malignant)
- 153.* Colon neoplasm (primary malignant)

Appendix 3
Liste des variables:

Demographic:

patid	Numéro	numéro d'identification du patient
cas	0/1	indicateur cas/contrôle (cas=1). Attention, étant donné l'appariement, il est impératif que chaque cas soit suivi de SON contrôle
datin	jj/mm/aaaa	date de l'interview
where	d/h	lieu de l'interview: domicile(d) ou hôpital(h)
birthd	jj/mm/aaaa	date de naissance
birthw	code postal (Belg) pays(sinon)	lieu de naissance
sex	m/f	homme (m) ou femme (f)
datdiag	aaaa	année de diagnostic
diag	RAEB/RA/RAEBT/ CMMOL	classification FAB
nbchildok	nombre	nombre d'enfants vivants en bonne santé
nbchildill	nombre	nombre d'enfants malades (antécédent de maladie hémato-lympho et auto-immune)
nbchilddead	nombre	nombre d'enfants décédés
karyo		Karyotype

residence:

patid	numéro	numéro d'identification du patient
princ1 2 3 4	1:ville (a=ouvert, b=bord, c=centre) 2:village (a=ouvert, c=centre) 3:paysan 4:forêt	résidence principale (si plusieurs résidences, prendre en compte la résidence majoritairement occupée pendant le dernier tiers de la période) par période 1 2 3 4
post1 2 3 4	code postal (Belg) 0000 (sinon)	code postal de la résidence principale (si plusieurs résidences, prendre en compte la résidence majoritairement occupée pendant le dernier tiers de la période) période 1 2 3 4
sec1 2 3 4	1:ville (a=ouvert, b=bord, c=centre) 2:village (a=ouvert, c=centre) 3:paysan 4:forêt	résidence secondaire (si plusieurs résidences, prendre en compte la résidence majoritairement occupée pendant le dernier tiers de la période) par période 1 2 3 4
trop1 2 3 4	0: non 1: occasionnel 2: habiter	environnement " tropical " durant la période 1 2 3 4

travail:

patid	Numéro	numéro d'identification du patient
agefin	age	âge de fin de formation
codefor	code1/code2	code1: 1,2,3,4,ou 5 selon le niveau d'éducation atteint (1 pour primaire, 5 pour supérieur de type long)/ code2: 0 ou 1 selon la réalisation(1) de cours supplémentaires (type cours du soir)
ilo1	???	code premier emploi
debut1	aaaa	année début premier emploi
fin1	aaaa	année fin premier emploi
ilo2	???	code deuxième emploi
debut2	aaaa	année début deuxième emploi
fin2	aaaa	année fin deuxième emploi
ilo3	???	code troisième emploi
debut3	aaaa	année début troisième emploi
fin3	aaaa	année fin troisième emploi
ilo4	???	code emploi 4
debut4	aaaa	année début emploi 4
fin4	aaaa	année fin emploi 4
ilo5	???	code emploi 5
debut5	aaaa	année début emploi 5
fin5	aaaa	année fin emploi 5
hobby1	???	code hobby1
debut1	aaaa	année début hobby 1
fin1	aaaa	année fin hobby 1
hobby2	???	code hobby2
debut2	aaaa	année début hobby 2
fin2	aaaa	année fin hobby 2
hobby3	???	code hobby3
debut3	aaaa	année début hobby 3
fin3	aaaa	année fin hobby 3

Envrt:

patid	numéro	numéro d'identification du patient
ferme1 2 3 4	durée en années	habitation dans une ferme pendant la période 1 2 3 4
unif1 2 3 4	"	habitation dans une maison unifamiliale période 1 2 3 4
apart1 2 3 4	"	habitation dans un appartement période 1 2 3 4
repos1 2 3 4	"	habitation dans une maison de repos période 1 2 3 4
oth1 2 3 4	"	habitation (autre) période 1 2 3 4
eaucour1 2 3 4	"	système d'alimentation en eau: eau courante période 1 2 3 4
src1 2 3 4	"	source période 1 2 3 4
cit1 2 3 4	"	citerne période 1 2 3 4
oth1 2 3 4	"	autre période 1 2 3 4
centmaz1 2 3 4	"	chauffage central au mazout période 1 2 3 4
centelec1 2 3 4	"	chauffage central électrique période 1 2 3 4
centgaz1 2 3 4	"	chauffage central au gaz période 1 2 3 4
charb1 2 3 4	"	poêle charbon/mazout/pétrole période 1 2 3 4
gaz1 2 3 4	"	poêle au gaz période 1 2 3 4
feu11 2 3 4	"	feu ouvert occasionnel période 1 2 3 4
feu21 2 3 4	"	feu ouvert qqs fois par mois période 1 2 3 4
feu31 2 3 4	"	feu ouvert qqs fois par semaine période 1 2 3 4
papier1 2 3 4	1:occasionnel 2: qqs fois par mois 3: qqs fois par sem.	code brûlage papier période 1 2 3 4
plastic1 2 3 4	"	code brûlage plastic période 1 2 3 4
bois1 2 3 4	"	code brûlage bois période 1 2 3 4
brul1 2 3 4	"	code brûlage autre période 1 2 3 4
inond1 2 3 4	nombre	nombre d'inondation pendant la période 1 2 3 4
plomb1 2 3 4	"	nombre de problèmes de plomberie pendant la période 1 2 3 4
domeau1 2 3 4	"	nombre d'autr(s) dommage(s) des eaux pendant la période 1 2 3 4
egout1 2 3 4	"	nombre de problèmes d'égouttage pendant la période 1 2 3 4
moisi1 2 3 4	"	nombre de problèmes de moisissure pendant la période 1 2 3 4
tach1 2 3 4	"	nombre de problèmes de tâches noires pendant la période 1 2 3 4
humid1 2 3 4	"	nombre de problèmes d'humidité pendant la période 1 2 3 4
fent1 2 3 4	o/f	fenêtres habituellement fermées (f) ou ouvertes (o) pendant la période 1 2 3 4
sdb1 2 3 4	0/15/30/60/+ 60	durée (min/jour) d'ouverture des fenêtres de la sdb pendant la période 1 2 3 4
chbre1 2 3 4	"	durée (min/jour) d'ouverture des fenêtres de la chbre pdt la période 1 2 3 4
odeur1 2 3 4	nombre	nombre de problèmes d'odeur d'humidité/moisissure période 1 2 3 4
net1 2 3 4	"	nombre de nettoyage consécutif à ces problèmes période 1 2 3 4
peint1	durée	durée en mois travaux de peinture période 1
dur1act	"	durée en mois travaux de peinture actif période 1
dur1pas	"	durée en mois travaux de peinture passif période1
peint2	"	2
dur2act	"	2
dur2pas	"	2
peint3	"	3
dur3pas	"	3
dur3act	"	3
peint4	"	4
dur4pas	"	4
dur4act	"	4
constr1	"	durée en mois travaux de construction période 1
dur1act	"	durée en mois travaux de construction actif période 1
dur1pas	"	durée en mois travaux de construction passif période 1
constr2	"	2
dur2act	"	2
dur2pas	"	2
constr3	"	3
dur3act	"	3

dur3pas	''		3
constr4	''		4
dur4act	''		4
dur4pas	''		4
restaur1	''	durée en semaine travaux de restauration	période 1
dur1act	''	durée en semaine travaux de restauration actif	période 1
dur1pas	''	durée en semaine travaux de restauration passif	période 1
restaur2	''		2
dur2act	''		2
dur2pas	''		2
restaur3	''		3
dur3act	''		3
dur3pas	''		3
restaur4	''		4
dur4act	''		4
dur4pas	''		4
peri-od1 2 3 4	0/1	période d'odeur de solvant remarquable (oui/non = 1/0)	période 1 2 3 4
ver1 2 3 4	nombre	nbre traitement chimique de l'habitation pour les vers du bois	période 1 2 3 4
humid1 2 3 4	''	nbre traitement chimique de l'habitation pour l'humidité	période 1 2 3 4
caf1 2 3 4	''	nbre de traitement cafard/rat	période 1 2 3 4
sitiindpol1 2 3 4	distance/-	distance en m. (- si rien) d'un site industriel polluant	période 1 2 3 4
sitindhopol1 2 3 4		distance en m. (-si rien) d'un site industriel non polluant	période 1 2 3 4
lht1 2 3 4		distance en m. (- si rien) d'un ligne haute tension	période1 2 3 4
centnuc1 2 3 4		distance en m. (- si rien) d'une centrale nucléaire	période 1 2 3 4
dech1 2 3 4		distance en m. (- si rien) d'une décharge	période 1 2 3 4
incin1 2 3 4		distance en m. (- si rien) d'une usine d'incinération par	période 1 2 3 4
hopi1 2 3 4		distance en m. (- si rien) d'un hôpital	période 1 2 3 4
aero1 2 3 4		distance en m. (- si rien) d'un aéroport	période 1 2 3 4
avion1 2 3 4	0/1	perturbation de la conversation par des avions oui/non (1/0)	période 1 2 3 4
kero1 2 3 4		odeur de kérosène oui/non (1/0)	période 1 2 3 4
traf1 2 3 4	distance/-	distance en m. (- si rien) d'une route à grand trafic	période 1 2 3 4
champ1 2 3 4		distance en m. (- si rien) de champs	période 1 2 3 4
foot1 2 3 4		distance en m (- si rien) de terrains de foot	période 1 2 3 4

Envrt trvl

patid	numéro	identification du patient
odeur1 2 3 4	nombre	nombre de problèmes d'odeur d'humidité/moisissure période 1 2 3 4
net1 2 3 4	"	nombre de nettoyage consécutif à ces problèmes période 1 2 3 4
peint1	durée	durée en mois travaux de peinture période 1
actif	"	durée en mois travaux de peinture actif période 1
passif	"	durée en mois travaux de peinture passif période1
peint2	"	2
actif	"	2
passif	"	2
peint3	"	3
actif	"	3
passif	"	3
peint4	"	4
actif	"	4
passif	"	4
constr1	"	durée en mois travaux de construction période 1
actif	"	durée en mois travaux de construction actif période 1
passif	"	durée en mois travaux de construction passif période 1
constr2	"	2
actif	"	2
passif	"	2
constr3	"	3
actif	"	3
passif	"	3
constr4	"	4
actif	"	4
passif	"	4
ver1 2 3 4	nombre	nbre traitement chimique du lieu de travail pour les vers du bois période 1 2 3 4
humid1 2 3 4	"	nbre traitement chimique du lieu de travail pour l'humidité période 1 2 3 4
caf1 2 3 4	"	nbre de traitement cafard/rat période 1 2 3 4
sitindpol1 2 3 4	distance/-	distance en m. (- si rien) d'un site industriel polluant période 1 2 3 4
sitindnopol1 2 3 4		distance en m. (-si rien) d'un site industriel non polluant période 1 2 3 4
lht1 2 3 4		distance en m. (- si rien) d'un ligne haute tension période1 2 3 4
centnuc1 2 3 4		distance en m. (- si rien) d'une centrale nucléaire période 1 2 3 4
dech1 2 3 4		distance en m. (- si rien) d'une décharge période 1 2 3 4
incin1 2 3 4		distance en m. (- si rien) d'une usine d'incinération période 1 2 3 4
hopi1 2 3 4		distance en m. (- si rien) d'un hôpital période 1 2 3 4
aero1 2 3 4		distance en m. (- si rien) d'un aéroport période 1 2 3 4
avion1 2 3 4	0/1	perturbation de la conversation par des avions oui/non (1/0) période 1 2 3 4
kero1 2 3 4		odeur de kérosène oui/non (1/0) période 1 2 3 4
traf1 2 3 4	distance/-	distance en m. (- si rien) d'une route à grand trafic période 1 2 3 4
champ1 2 3 4		distance en m. (- si rien) de champs période 1 2 3 4
foot1 2 3 4		distance en m (- si rien) de terrains de foot période 1 2 3 4

maladies:

patid	numéro	identification du patient
infect	0/1	présence(1)/absence (0) d'infection
year	aaaa	année de la première infection
hepatit	0/ABCY	absence d'hépatite ou catégorie d'hépatite
year	aaaa	première année de diagnostic
malaria	1/0	présence (1)/ absence (0) de malaria
year	aaaa	année de la première crise
tuberc	1/0	présence (1)/absence (0) de la tuberculose
year	aaaa	année début tuberculose
bronch	1/0	présence/absence (1/0) de bronchite chronique
year	aaaa	année première bronchite
allerg	1/0	présence/absence (1/0) allergie
year	aaaa	année première manifestation allergique
digest	1/0	présence/absence (1/0) de troubles digestifs
year	aaaa	année premier trouble digestif
cardio	1/0	présence/absence (1/0) de trouble cardiovasculaire
year	aaaa	année premier trouble cardiovasculaire
urin	1/0	présence/absence (1/0) de trouble urinaire
year	aaaa	année premier trouble urinaire
endoc	1/0	présence/absence (1/0) de trouble endocrinien
year	aaaa	année premier trouble endocrinien
os	1/0	présence/absence (1/0) d'affections articulaire/osseuse
year	aaaa	année première affection articulaire/osseuse
peau	1/0	présence/absence (1/0) d'affections de la peau
year	aaaa	année première affection de la peau
hemo	1/0	présence/absence (1/0) d'hémopathies
year	aaaa	année première hémopathie
anem	1/0	présence/absence (1/0) anémie
year	aaaa	année diagnostic première anémie
cancer	1/0	présence/absence (1/0) de tumeur/cancer
year	aaaa	année première tumeur:cancer
oth	1/0	présence/absence (1/0) d'autres maladies
year	aaaa	année première autre maladie
epilep	1/0	présence/absence (1/0) d'épilepsie
year	aaaa	année première épilepsie
migr	1/0	présence/absence (1/0) de migraine
year	aaaa	année première migraine
oper1/2/3/4	nombres	nombre opérations par période 1/2/3/4
depression	1/0	présence/absence (1/0) d'une dépression
year	aaaa	année de la première dépression

radio:

Les radios précédant la période 1 sont cumulées avec celles de la période 1

patid	numéro	identification
rxpoum 1 2 3 4	nombre	radiographie pulmonaire périodes 1, 2, 3, 4
scopoum 1 2 3 4	nombre	scopie pulmonaire périodes 1, 2, 3, 4
tompoum 1 2 3 4	nombre	tomographie pulmonaire périodes 1, 2, 3, 4
stom 1 2 3 4	nombre	radiographie oesophage/estomac périodes 1, 2, 3, 4
vesic 1 2 3 4	nombre	radiographie vésicule périodes 1, 2, 3, 4
baryt 1 2 3 4	nombre	lavement baryté (intestin) périodes 1, 2, 3, 4
rein1 2 3 4	nombre	radiographie rein périodes 1, 2, 3, 4
angio 1 2 3 4	nombre	radiographie pulmonaire périodes 1, 2, 3, 4
bassin 1 2 3 4	nombre	radiographie bassin/colonne périodes 1, 2, 3, 4
os1 2 3 4	nombre	radiographie autres os périodes 1, 2, 3, 4
dent1 2 3 4	nombre	radiographie une dent périodes 1, 2, 3, 4
dentt 1 2 3 4	nombre	radiographie dentaire panoramique périodes 1, 2, 3, 4
scancer 1 2 3 4	nombre	scanner cérébral périodes 1, 2, 3, 4
scanthor 1 2 3 4	nombre	scanner thoracique périodes 1, 2, 3, 4
scanrach 1 2 3 4	nombre	scanner rachidien périodes 1, 2, 3, 4
scanabdo 1 2 3 4	nombre	scanner abdominal périodes 1, 2, 3, 4
scanoth 1 2 3 4	nombre	scanner autre région périodes 1, 2, 3, 4
rmn 1 2 3 4	nombre	RMN cérébral/rachidien/ ... périodes 1, 2, 3, 4
scintpoum 1 2 3 4	nombre	scintigraphie poumon périodes 1, 2, 3, 4
scintos 1 2 3 4	nombre	scintigraphie des os périodes 1, 2, 3, 4
scinthyr 1 2 3 4	nombre	scintigraphie thyroïde périodes 1, 2, 3, 4
scintcoeur 1 2 3 4	nombre	scintigraphie du cœur périodes 1, 2, 3, 4
scinthoth 1 2 3 4	nombre	scintigraphie autre région périodes 1, 2, 3, 4
mamo1 2 3 4	nombre	mammographie périodes 1, 2, 3, 4
densitométrie	nombre	nombre d'examen densitométrique par période 1, 2, 3, 4

drugs:

patid	numéro	identification patient
medldurj	nombre	nombre d'année d'utilisation de la médication I durant la période J
medldurj	nombre	nombre de jour par an pendant lesquels la médication a été utilisée

toxic	1/0	indicateur de toxicologie
chimio1	1/0	traitement chimiothérapique période 1
deb1	aaaa	année debut traitement période 1
fin1	aaaa	année fin traitement période 1
chimio2	1/0	2
deb2	aaaa	2
fin2	aaaa	2
chimio3	1/0	3
deb3	aaaa	3
fin3	aaaa	3
chimio4	1/0	4
deb4	aaaa	4
fin4	aaaa	4
chimioct1	1/0	contact (oui=1) avec personne en traitement chimiotherapique periode 1

chimioct2	"	2
chimioct3	"	3
chimioct4	"	4

med1: phénilbutazone
 med2: cortisone
 med3: indométacine
 med4: immunosuppresseurs
 med5: chloramphénicol
 med6:
 med7:
 med8:
 med9:
 med10:

bmi:

patid	numéro	identification patient
taille	nombre	taille en cm
poids1	nombre	poids (kg) période 1
poids2	nombre	poids (kg) période 2
poids3	nombre	poids (kg) période 3
poids 4	nombre	poids (kg) période 4
poidsmin	nombre	poids min (kg)
poids max	nombre	poids max (kg)
maigr	0/1	1 si amaigrissement important derniere periode/ 0 sinon

famille:

Les cancers concernés par cette partie sont uniquement d'ordre hémato-lympho

patid	numéro	identification patient
percanc	0/1	cancer père (oui = 1)
mercanc	0/1	cancer mère
frere	nombre	nombre de frères
frercanc	nombre	nombre de frères avec cancer
soeur	nombre	nombre de soeurs
soeurcanc	nombre	nombre de soeurs avec cancer
fiils	nombre	nombre de fils
fiilscanc	nombre	nombre de fils avec cancer
filie	nombre	nombre de fille
fiilcanc	nombre	nombre de fille avec cancer

alimentation:

patid	numéro	identification patient
cafe 1 2 3 4	nombre	nombre de tasse(s) de cafe par jour * 1(léger), 2 (moyen) 3 (fort) par période 1,2,3,4
the 1 2 3 4	nombre	nombre de tasse(s) de thé par jour par période 1 2 3 4
cola 1 2 3 4	nombre	nombre de verre(s) par jour par période 1 2 3 4
limo 1 2 3 4	nombre	nombre de verres de limonade par période 1 2 3 4
lait 1 2 3 4	nombre	nombre de tasse(s) par jour de lait frais par période 1 2 3 4
vin 1 2 3 4	nombre	nombre de verre(s) vin par jour en semaine par période 1 2 3 4
vinWE 1 2 3 4	nombre	nombre de verre(s) vin par jour en WE par période 1 2 3 4
apero 1 2 3 4	nombre	nombre de verre(s) apéritif par jour en semaine par période 1 2 3 4
aperoWE 1 2 3 4	"	nombre de verre(s) apéritif par jour en WE par période 1 2 3 4
biere 1 2 3 4	nombre	nombre de verre(s) biere par jour en semaine par période 1 2 3 4
biereWE 1 2 3 4	"	nombre de verre(s) biere par jour en WE par période 1 2 3 4

Tabac:

patid	numero	Identification
tab1 à tab40	0:jamais 1:passif(occasion.) 2:passif(beaucoup) 3:occasionnel 4:sortie ou repas 5:avec filtre < 1/2 pqt /jour 6:sans fltre < 1/2 pqt/jour 7:avec filtre [1/2-1]pqt/jour 8:sans filtre [1/2-1]pqt/jour 9: > 1 pqt/jour 10: > 2 pqt/jour	code consommation tabac pour chaque année
pipe	nombre	nombre d'années avec consommation de pipe par période
cigare	nombre	nombre d'années avec consommation de cigare par période

pesticide:

patid	numéro	Identification
ferme 1 2 3 4	nombre	distance (m.) d'une ferme par période 1 2 3 4
champ 1 2 3 4	nombre	distance (m.) de champs par période 1 2 3 4

jard 1 2 3 4	nombre	distance (m.) d'un lieu où on utilise des pests. par période 1 2 3 4
stock 1 2 3 4	nombre	distance (m.) stockage industriel de pdts alim. par période 1 2 3 4
bois 1 2 3 4	nombre	distance (m.) stockage industriel de bois par période 1 2 3 4
flor1 2 3 4	nombre	distance (m.) floriculture ou serres par période 1 2 3 4
bet 1 2 3 4	nombre	distance (m.) d'un éleveur de bétail par période 1 2 3 4
pestIJ	nombre	durée en année(s) d'utilisation de fongicide/insecticide/herbicide pendant la période J (1 2 3 4) au niveau I (1(occasionnel), 2(qqfs par mois) ou 3 (qqfs par semaine))
ddtIJ	nombre	nombre d'année(s) concernée par l'utilisation de DDT pendant la période J (1 2 3 4) au niveau I (1(occasionnel), 2(qqfs par mois) ou 3 (qqfs par semaine))
oth 1,2,3,4	nombre	nombre d'années concernées par l'exposition à un autre pesticide durant les périodes 1, 2, 3, 4 (hélicoptère)
cod1 2 3 4	code	code d'exposition aux pesticides pendant les périodes 1 2 3 4

solvants:

patid	numero	identification du patient
trichlo 1 2 3 4	nombre	nombre d'année(s) concernée(s) par l'utilisation de trichloroéthylène pendant les périodes 1 2 3 4
tolu 1 2 3 4	nombre	nombre d'année(s) concernée(s) par l'utilisation de toluol xylo pendant les périodes 1 2 3 4
napht 1 2 3 4	nombre	nombre d'année(s) concernée(s) par l'utilisation de naphte pendant les périodes 1 2 3 4
decap 1 2 3 4	nombre	nombre d'année(s) concernée(s) par l'utilisation de naphte pendant les périodes 1 2 3 4
mastic 1 2 3 4	nombre	nombre d'année(s) concernée(s) par l'utilisation de colles mastiques pendant les périodes 1 2 3 4
ammo 1 2 3 4	nombre	nombre d'année(s) concernée(s) par l'utilisation d'ammoniaque pendant les périodes 1 2 3 4
clor1 2 3 4	nombre	nombre d'année(s) concernée(s) par l'utilisation de produit chloré pendant les périodes 1 2 3 4
typexpo 1 2 3 4	code	type d'exposition pendant les périodes 1 2 3 4
freqexpo 1 2 3 4	code	fréquence d'exposition pendant les périodes 1 2 3 4

Divers:

patid	numero	Identification
hormone	code 1/0/NA	manipulation d'hormone (oui=1, non= 0, ne sait pas = NA)
huile	code 1/0/NA	manipulation d'huile de transfo/condensateur (oui=1, non=0, NA sinon)
lht	code 1/0/NA	environnement professionnel de lignes à haute tension (oui=1, non =0, NA sinon)
bronz	code 1/2/3/NA	description de la réaction de la peau au soleil: 1= " bronze sans coup de soleil ", 2= " bronze après coup de soleil ", 3= " continue à rougir ", NA= " ne sait pas "

dioxine:

patid	numero	Identification
batim 1 2 3 4	nombre	nombre d'incendie(s) de bâtiment par période 1 2 3 4
champ 1 2 3 4	nombre	nombre d'incendie(s) de champ par période 1 2 3 4
foret 1 2 3 4	nombre	nombre d'incendie(s) de forêts par période 1 2 3 4
bois 1 2 3 4	nombre	nombre de feu(x) en plein air avec du bois par période 1 2 3 4
caout 1 2 3 4	nombre	nombre de feu(x) en plein air avec du caoutchouc par période 1 2 3 4
chim 1 2 3 4	nombre	nombre de feu(x) en plein air avec des produits chimiques par période 1 2 3 4
plastic 1 2 3 4	nombre	nombre de feu(x) en plein air avec plastiques par période 1 2 3 4
cod 1 2 3 4	code	code d'exposition à la dioxine par période

Appendix 4: Some abstracting instructions
--

Residence:

For each decade, in case of multiple residences used as principal residence, the last occupied residence will be considered as long as it has been used for more than 3 years. If this is not the case, the residence occupied for the longest time-period will be considered, with higher priority given to more recent residences in case of equality. The rationale for this choice is the belief that more recent residences might probably have a bigger impact than older residence.

Occupation

Professional occupations will be coded using ISCO-68 codes (three number codes for minor and unit groups) and NACE codes (four numbers coding based upon the 1993 « Nomenclature des Activités économiques dans les Communautés Européennes).

Medication

All drugs encountered will be classified in categories which will reflect the questioning experience, and for each category, and by decade, the number of years during which the drug was used will be recorded together with the mean number of days of use per year.

Annex V

MDS case-control study:

General registration document for case and referent interviews

Investigators / Authors:

Dr Karel Van Damme

Dr. Ludwine Casteleyn

Dr. Alix Collard

Annex VI

MDS case-control study: Examples of the specific questionnaires

Investigators / Authors:

Dr. Karel Van Damme

Dr. Ludwine Casteleyn

Dr. Alix Collard

Dr. Lucia Miligi

Prof. Dr. Paolo Vineis

Annex VII

MDS case-control study: Statistical analysis report

Contents:

1. Introduction
2. Objectives
3. Methods
4. Results

Investigators / Authors:

Dr Karel Van Damme

Dr. Ludwine Casteleyn

Dr. Alix Collard

1. Introduction

See study protocol (annex IV)

2. Objectives

See study protocol (annex IV)

3. Methods

3.1. Study population

For this case-control study, all cases of MDS (according to French-American-British (FAB) diagnostic criteria) diagnosed between 1/1/1997 and 31/01/1999 at St Luc or Gasthuisberg hospital, aged between 20 and 75 years are included, with the exception of MDS cases secondary to antineoplastic agents and patients who were unable to participate during the interviewing period.

For every case of MDS, several possible reference cases were selected based upon the medical archives from the same hospital as the case's, with one of the following pathologies: renal cell cancer, colon cancer or gastro-duodenic ulcer, matched for age (five years interval) and sex. For each of the three reference pathologies, the last two persons who were either hospitalised, or seen at the consultation were selected from the medical archives, and ranked in a random fashion. These referent-patients were then contacted in that order until someone agreed to be interviewed.

3.2. Rating instruments

3.2.1. Mode of interview

The same two occupational health physicians interviewed all cases and controls using a face-to-face interview, taking the form of a structured conversational approach based upon a structured registration document or questionnaire. The information collected included demographic data, lifetime residence, occupational history and hobbies, medical history including drugs, X-ray procedures, radiotherapy, operations, past illnesses, and life style data like smoking and drinking habits or diet. For all the information collected, the period concerned was recorded on a time scale which covered the four last decades preceding diagnosis

3.2.2. Scales

The living environment is described through one of the following codes for each decade:

- 1a for town-open space,
- 1b for town-periphery,
- 1c for city-centre,
- 2a for village-open space,
- 2b for village-centre,
- 3 for countryside,
- 4 for forest.

For each decade, possible contact with a tropical environment will also be coded through one of the following codes: 0 for no contact, 1 for occasional contact, 2 for residing for longer periods. Education and training level is coded as described in table 1.

Table 1: Codes for education and training level.

General education		Supplementary training	
Primary school	code 1	Yes	code 1
Secondary school	code 2	No	code 0
Professional/Technical training	code 3		
Superior short-type	code 4		
Superior long-type	code 5		

Coffee consumption is expressed in cups of light coffee using the convention that one cup of strong coffee is considered to correspond to 3 cups of light coffee, and one cup of normal coffee to correspond to 2 cups of light coffee.

Tobacco consumption is described through an ordinate scale going from « never », code 0 up to 10 for a daily consumption of more than two packets, for each of the forty years preceding the diagnosis.

History of exposure to pesticides and of drug consumption will also be coded by experts blinded to the case/control status, starting from the original data.

3.3 Data management

Whereas the questionnaire is designed to be useful during the process of collecting information, the database is designed to translate this collected information in an efficient way into a data structure which can be exploited for statistical analysis. Hence, all the information collected during the interview is translated into a series of variables, either of character type, or using numerical formats or date formats (see protocol for the list of variables and associated codes and formats). For most of the information collected, such as birth date, sex, ... the information collected during the interview can be directly translated into the corresponding variables. However, some pre-coding work was necessary, for instance in case multiple codes had been used for the same period. In these cases, the majority (of occurrence) rule is applied, and in case of equality the code corresponding to the sub-period closest to the time of diagnosis is given for the whole period.

3.4. Statistics

Usual summary statistics will be provided for description of variables. Conditional logistic regression will be used to compare cases and controls and to provide relative risk estimates and associated 95% confidence interval.

4. Results

4.1. Initial characteristics of the case and referent population.

This report concerns 30 pairs of MDS patients with their matched control, among which 8 are from the renal cell cancer pathology group, 8 from the colon cancer pathology group and 12 from the gastro-duodenic ulcer group, and 2 have other pathologies (hernia and polykystic kidney) but were mentioned in the hospital database as presenting one of the three target control pathologies. These patients will all be included for the analysis, except the control patient (and hence his matched case) with polykystic kidneys who will be excluded for the drug consumption analysis.

Median age of the cases was 68.5 years (range 52-79). Matching criteria were respected for all pairs except one (patients 25 and 26 for whom the age difference is 6 years and 5 months, control category: kidney). This pair is kept in the analysis. There were 23 (77%) males and 7 (23%) females. According to FAB classification, there were 4 cases of refractory anaemia (RA), 1 case of refractory anaemia with ringed sideroblasts (RARS), 9 cases of refractory anaemia with excess of blasts (RAEB), 5 cases of RAEB in transformation (RAEB-T) and 3 cases of chronic myelomonocytic leukaemia (CMML). For 8 patients, we have no data on MDS subgroup. Median age of the control group is 67.5 years (range 50-82). The age range accepted in the study has thus been extended to include 6 controls and 4 cases aged more than 75 years, in order to be able to reach a reasonable sample size.

As can be seen from table 1, the final training level of cases is significantly lower in cases than in controls, and this observed difference is statistically significant at the usual 5% level (p -value = 0.0280).

Table 1: Socio-economic characteristics of the MDS and control populations:

	MDS cases	Controls	P-value
Sex			
Male	23	23	
Female	7	7	
Age (mean \pm SD)	66.33 \pm 7.67	66.70 \pm 8.48	0.8649(Wilcoxon test)
Education level			
Primary school	8	5	
Secondary school	14	8	
Professional/ technical	5	8	
Superior short type	3	4	
Superior long type	0	5	0.0280

4.2. Demographic data

Twenty-eight (28) cases and 24 controls had at least one child. The mean number of children in those persons was 2.60 ± 2.11 in MDS cases and 2.23 ± 1.98 in controls (p-value for comparison: 0.5185).

Table 2: demographic characteristics of the MDS and control populations.

	MDS cases	Controls	P-value for comparison
Number of children			
0	2	6	
1	3	3	
2	14	10	
3	7	7	
4	1	2	
5	2	1	
More	1	1	0.5183
Familial history of cancer: number of cancers			
0	14	9	
1	10	10	
2	1	5	
3	1	0	
4	0	2	
5	0	1	
6	1	0	0.1999

The familial history of haematolymphopoi etic cancer was also studied through a score computed as the sum of

- 1 if such a cancer is reported for the patient's father
- 1 if such a cancer is reported for the patient's mother
- number of brother with such a cancer / total number of brother (0 if no brother)
- similar computations for sister, son and daughter

The score is not available for 3 control patients (patient identification number 8, 50, 60). The pairs of patients corresponding to these controls is thus not included in this analysis. The derived score have been compared across cases and controls and no statistically significant difference was detected (p-value = 0.1999). The mean score for the cases is 0.58 while it is 0.89 for the controls.

Concerning primary residence, patients main characteristics are summarised in table 3. Grouping categories 1a, 2a, 3 and 4 versus categories 1b, 1c and 2c, in order to compare « crowded » versus « quiet » environment did not show any difference between cases and controls, neither did the comparison between « pesticide » versus « clean » environment using group « 1a, 2a, 3 » versus « 1b, 1c, 2c, 4 » (see table 4).

Table 3: Type of primary residence per period for cases and controls.

Type of res.	1a	1b	1c	2a	2c	3	4
Period 1							
Number of controls	3	8	5	10	0	2	2
Number of cases	7	6	6	8	2	1	0
Period 2							
Number of controls	4	6	4	11	1	3	1
Number of cases	6	7	3	10	2	2	0
Period 3							
Number of controls	4	5	2	13	2	3	1
Number of cases	5	6	2	12	3	2	0
Period 4							
Number of controls	4	5	2	13	2	3	1
Number of cases	5	6	2	12	3	2	0

Table 4: Comparison of residential environment between cases and controls.

« crowded » versus « quiet » environment i.e. group « 1a, 2a, 3,4 » vs. « 1b, 1c, 2c »

Crowded	period1		period2		period3		period4	
	yes	no	yes	no	yes	no	yes	no
Cases	14	16	12	18	11	19	11	19
Controls	13	17	11	19	9	21	9	21

« pesticide » versus « clean » environment: i.e. group « 1a, 2a, 3 » Vs « 1b, 1c, 2c, 4 »

Pesticide	period1		period2		period3		period4	
	yes	no	yes	no	yes	no	yes	no
Cases	16	14	18	12	19	11	19	11
Controls	15	15	18	12	20	10	20	10

4.3. Tobacco consumption

The mean tobacco score has been computed for each period and is described in the following table for cases and controls.

Table 5: Tobacco consumption for cases and controls: summary of the distribution characteristics for the mean score level within cases and controls for each decade.

	Period4	Period3	Period2	Period1
Cases				
Mean ± sd	4.84 ± 3.73	4.91 ± 3.39	3.18 ± 3.34	2.14 ± 3.29
Min - med - max	0 - 6 - 10	0 - 6 - 10	0 - 2.2 - 10	0 - 4.2 - 10
Controls				
Mean ± sd	3.68 ± 3.86	3.93 ± 3.90	3.55 ± 3.67	2.86 ± 3.70
Min - med - max	0 - 2 - 10	0 - 2 - 10	0 - 2.2 - 10	0 - 0 - 10

If we dichotomise the tobacco scores between non-smokers (i.e. people who never smoked, people reporting passive smoking and people reporting smoking only very rarely, during «parties » ..., scores 1,2,3 or 4) and smokers (score > 4, whatever the consumption on a daily basis), the effect of having ever smoked appears significant for the fourth period (between 40 and 30 years before diagnosis (p-value = 0.0461), with a relative risk estimated to be 3.33 (associated confidence interval: [1.00, 12.09]), but not in the subsequent periods. Classification of cases and controls in smokers and non-smokers for each period is provided in the table 6.

Table 6: classification of case and controls in smoker/non-smoker category for each period.

	Period4		Period3		Period2		Period1	
Smokers	yes	no	yes	no	yes	no	yes	no
Cases	19	11	19	11	10	20	9	21
Controls	12	18	13	17	10	20	8	22
P-value for comparison*	0.0461		0.1031		NS		NS	

*based on conditional logistic regression

4.4. Coffee consumption

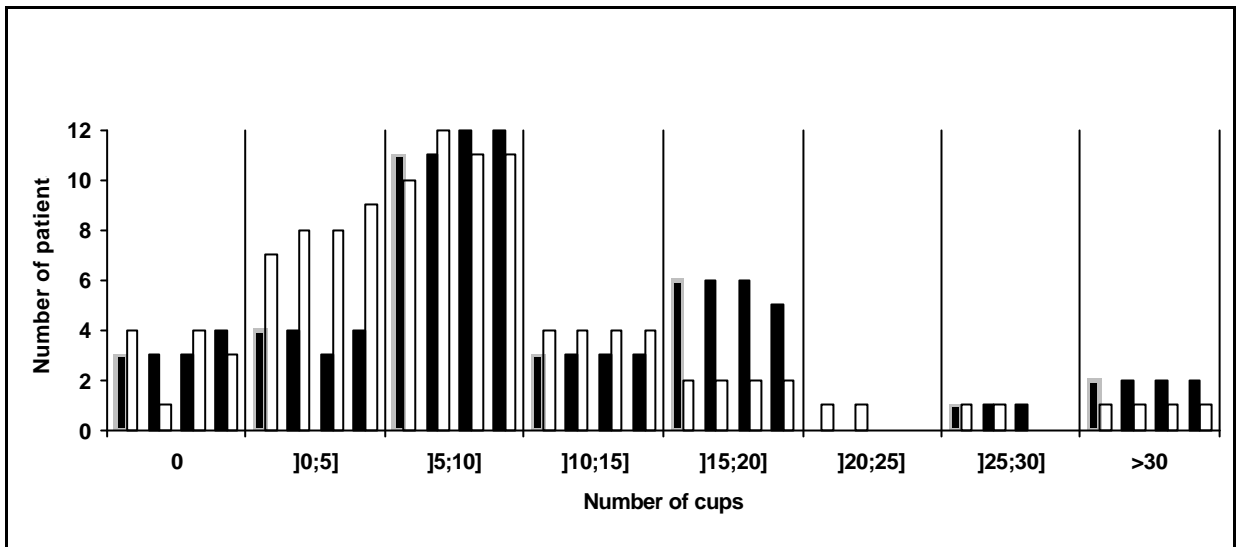
The mean coffee consumption by period and separately for cases and controls is given in table, and although the mean level of consumption is always more important in cases than in controls, no statistically significant difference was detected between cases and controls. The comparison between cases and controls remained non significant when considering the cumulative consumption.

Table 7: Mean [min; max]coffee consumption in cases and controls, for each period

Period	4	3	2	1
Cases	12.57 [0;60]	12.57 [0;60]	12.7 [0;60]	11.03 [0;60]
Controls	9.47 [0;45]	10.03 [0;45]	7.8 [0;45]	8 [0;45]
p-value*	NS	NS	0.1004	NS

* P-value for the comparison of cases and controls using conditional logistic regression

Figure 1: Coffee consumption. Within each consumption category, the four successive periods are depicted, with plain black boxes corresponding to cases and empty boxes to controls.



4.5. Medication:

Medications are described using a list of 63 categories described in annex. Patients 19 and 20 are excluded from this analysis, since patient 20 had a polykystic kidney diagnosis, which is thought to influence the medication consumption description.

Main results concerning drug consumption are described in table 8. The medications described as not being associated with the pathologies of interest are more frequently observed within the control group. The Zyloric consumption appears to be more frequent in the case group during the last period. The list of suspect medications also appears to correspond to a significant risk factor, all the more when the antibiotics are included in it. The antibiotics show a highly significant positive association with the case status, through all periods.

Table 8: The distribution of cases and controls have been compared for different medication categories determined by the experts, who have thus provided target lists of specific medications. For each list, the patients classification with respect to consumption of one of the medications included in the target list and case/control status is presented for the period indicated in the column header, and the p-values for the comparison of case versus control (conditional logistic regression). NS indicates non-significant comparison.

	Period 1		Periods 1,2		Ever/never	
	Yes	no	yes	no	yes	No
List of medications not associated with the pathologies of interest (4,5,6,14,15,17,21,22,24,27,31,34,35,36,37,44,47,48,51,52,53,54,56,60,61) [Note that the effect is more of this type of medications in the control group]						
Case	18	11	18	11	18	11
Control	22	7	23	6	23	6
p-value	0.0975		0.0530		0.0530	
GI-Ulcus (1)						
	yes	no	yes	no	yes	no
Case	5	24	6	23	6	23
Control	4	25	4	25	6	23
p-value	0.2077		0.1506		0.2390	
Zyloric (2)						
	Yes	no	yes	no	yes	no
Case	3	26	3	26	4	25
Control	1	28	1	28	1	28
p-value	0.1188		0.1188		0.0687	
Anti-malaria (16)						
	Yes	no	yes	no	yes	no
Case	1	28	2	27	4	25
Control	1	28	2	27	3	26
p-value	NS		NS		0.2162	
List of suspect medications (without antibiotics) (12,13,40,62,20,11,2,10,19,20,23,45,49,55,59)						
	Yes	no	Yes	no	yes	no
Case	14	15	15	14	16	13
Control	8	21	8	21	11	18
p-value	0.0547		0.0369		0.0992	

List of suspect medications (including antibiotics) (12,13,40,62,20,11,2,10,19,20,23,45,49,55,59,3,52,33)						
	Yes	no	yes	no	yes	no
Case	18	11	20	9	21	8
Control	9	20	9	20	12	17
p-value	0.0114		0.0043		0.0205	
Antibiotics (3)						
	Yes	no	yes	no	yes	no
Case	6	23	8	21	9	20
Control	1	28	1	28	1	28
p-value	0.0039		0.0009		0.0004	
Alkylants and antimetabolites (45 & 55)						
	Yes	No	yes	no	yes	no
Case	2	27	2	27	2	27
Control	0	29	0	29	0	29
p-value	0.0414		0.0414		0.0414	
List 19 antireuma + 20indometacine + 32 tanderil + 11 nsaif						
	Yes	No	yes	no	yes	no
Case	9	20	9	20	11	18
Control	3	26	3	26	5	24
p-value	0.0334		0.0334		0.0444	
List 19 antireuma + 20 indometacine						
	Yes	No	yes	no	yes	no
Case	0	29	1	28	2	27
Control	0	29	0	29	1	28
p-value	NS		0.0960		0.1889	
List 32 tanderil + 11 nsaif						
	Yes	No	yes	no	yes	no
Case	9	20	9	20	10	19
Control	3	26	3	26	4	25
p-value	0.0334		0.0334		0.0334	
Cortisone (23)						
	Yes	No	yes	no	yes	no
Case	3	26	4	25	4	25
Control	3	26	3	26	3	26
p-value	NS		0.2162		0.1810	

List other hormones: 12 13 62 40						
	Yes	no	yes	no	yes	no
Case	5	24	5	24	5	24
Control	3	26	3	26	3	26
p-value	0.1506		0.1506		0.1506	

A supplementary analysis has been performed, based upon the following score defined by the medical experts:

- 1 : No medication mentioned for the decade preceding the diagnosis
- 2 : Regular intake during last decade before diagnosis of medications not in the list of medications included in the definition of higher score levels
- 3 : Occasional intake during last decade before diagnosis of NSAIF (1) or more « frequent » intake of these medications but only during first part of the forty years
- 4 : Not defined intermediate between 3 and 5
- 5 : regular intake of NSAIF and/or hormonal products during the decade preceding diagnosis
- 6 : Intake of drugs suspected to act on the haematolymphopoietic system during the decade preceding diagnosis (period 1)

Patients classification is provided in table 9. Comparison of score levels across cases and controls leads to a p-value of 0.0921, with an estimated RR = 1.32, and associated 95 % confidence interval [0.94, 1.87]. Comparison of category 6 versus all others leads to a p-value of 0.0465, a point estimate for the relative risk of RR = 6.00, with associated 95% confidence interval [1, 48.36]. Comparison of categories 5 and 6 versus all others leads to a p-value of 0.1601, a point estimate for the relative risk of RR = 2.25, and associated 95% confidence interval [0.69; 7.31]. Comparison of categories 3,4,5 and 6 versus 1 and 2 leads to a p-value of 0.0162, a point estimate for the relative risk RR = 4.00, and associated 95% confidence interval [1.13; 14.12].

Table 9: Cases and control distributions across drug score categories.

Medication score	1	2	3	5	6
Cases	5	7	6	5	6
Controls	2	19	2	5	1

4.6. Pesticides:

Pesticide exposure coming from any type of source (professional, environmental, ...) has been examined by a group of experts blinded to the case/control status, and a score attributed to each patient according to the following scale:

0 = no mentioning of any exposure

- 1 = sporadic exposure plausible but not certain, and no complaint of any adverse effect
- 2 = sporadic use reported by the patient to be very occasional
- 3 = sporadic passive exposure due to a generous pesticide use by others, and no complaint of any adverse effect
- 4 = from time to time passive exposure, but reported as a nuisance, or active use by the patient reported to be more frequent, although still occasional
- 5 = frequent passive exposure linked with seasons, and felt as a nuisance
- 7 = intense exposure clearly felt as a nuisance, or involving a direct contact with the product, or active exposure with a frequency higher than that usually observed
- 9 = clear passive or active exposure, almost on a daily basis, without any protection

Some intermediate categories were not defined by the experts, but were felt to be necessary in order to be able to include intermediate exposure histories which could have occurred, but were not observed in this sample. The distribution of cases and controls into these categories is described in table 10.

Table 10: Classification of cases and controls according to their score on the pesticide scale (see text for details), for each decade.

Code	0	1	2	3	4	5	7	9
Period 4								
Cases	16	1	6	1	2	1	1	2
Controls	20	2	4	0	1	1	1	1
Period 3								
Cases	11	1	5	1	6	1	3	2
Controls	18	2	5	0	3	0	1	1
Period 2								
Cases	8	2	3	1	9	2	3	2
Controls	16	3	5	0	3	0	2	1
Period 1								
Cases	8	2	3	1	10	1	4	1
Controls	15	3	6	0	3	0	2	1

In order to compare the pesticide exposure levels between cases and controls, single degree of freedom tests for a trend in relative risk with increasing level of exposure were applied to the observed score levels, for the cumulative exposure, and for each period:

- Having computed the cumulative score for all patients, which should reflect the lifetime exposure, the following table indicates that the mean level of exposure is more important in cases. In order to further explore this difference, the observed score values were compared between cases and controls taking the matching design into account. The resulting relative risk estimate is 1.08 (significantly different from 1, p-value = 0.02) with confidence interval [1.00; 1.17]. This relative risk represents the estimated increased risk for a change of one unit in exposure level. In other words, the relative risk is fixed at one for the category with code 0, and the relative risk is 1.08 for belonging to category 1, $1.08^2 = 1.17$ for belonging to category 2, $1.08^5 = 1.47$ for belonging to category 5 and $1.08^9 = 2.00$ for belonging to category 9.

Table 11: Summary distribution of cumulative pesticide exposure score over 4 decades, within case and control groups.

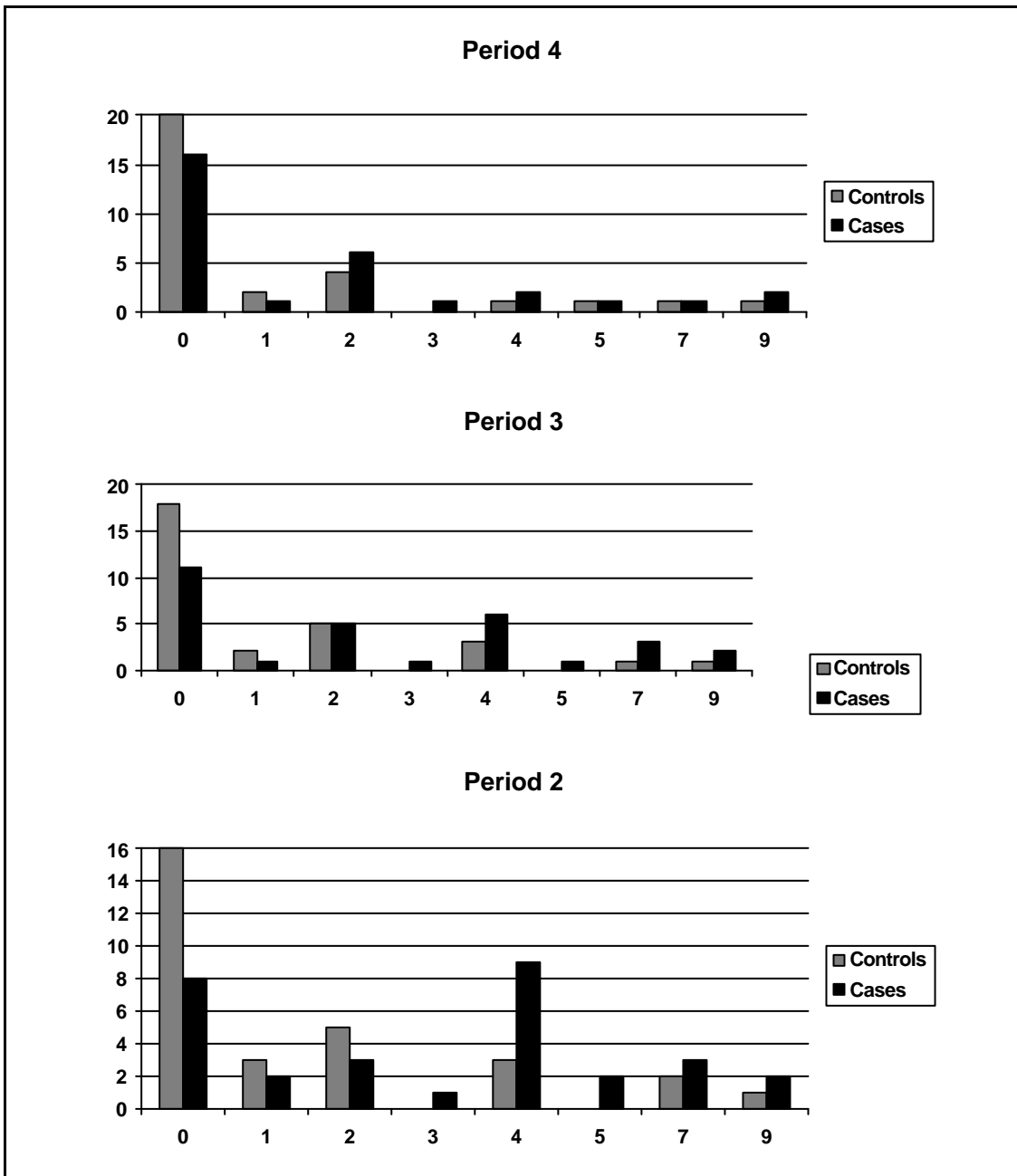
	Cases	Controls	P-value
Number of observations	30	30	
Mean \pm sd	10.83 \pm 9.45	5.67 \pm 8.95	
Min - med - max	0 - 11.5 - 36	0 - 1 - 36	0.02

- The scores have also been compared separately for each period between cases and controls. For period 4, although the observed mean level is higher (1.80) in the cases than in the controls (1.17), the observed difference did not reach statistical significance. However, for periods 3, 2 and 1, the difference between the two groups increase, and the statistical significance is reached. Corresponding relative risks and associated confidence intervals for an increase of one unit in pesticide exposure level code is provided in table 12.

Table 12: Relative risk and associated confidence interval for an increase of one unit in the pesticide exposure score attributed by expert, by decade.

Relative risk	Period 4	Period 3	Period 2	Period 1
Point estimate	1.12	1.32	1.42	1.37
Confidence interval	[0.89; 1.41]	[1.01;1.72]	[1.04; 1.90]	[1.04; 1.82]

Figure 2: Histogram for pesticide exposure level, for each period



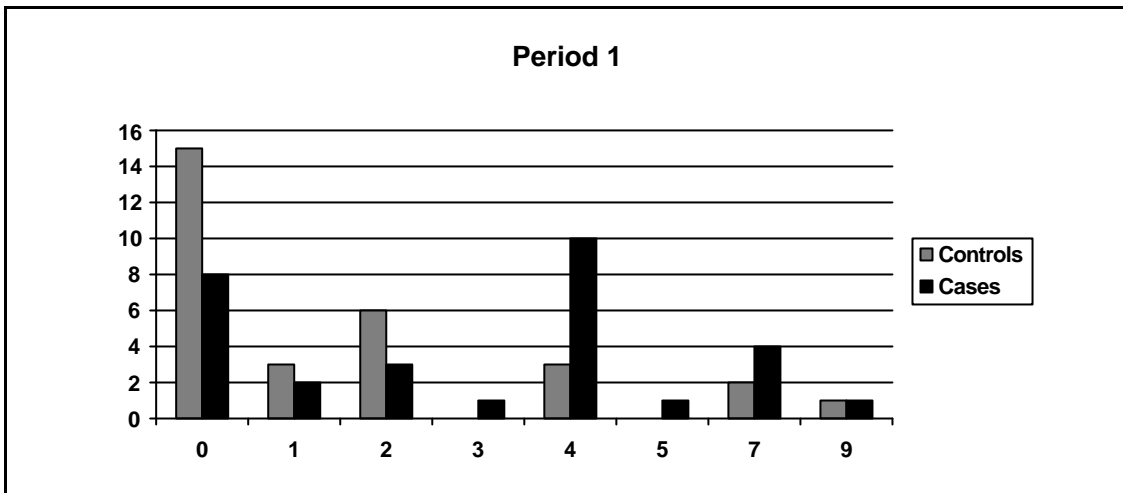
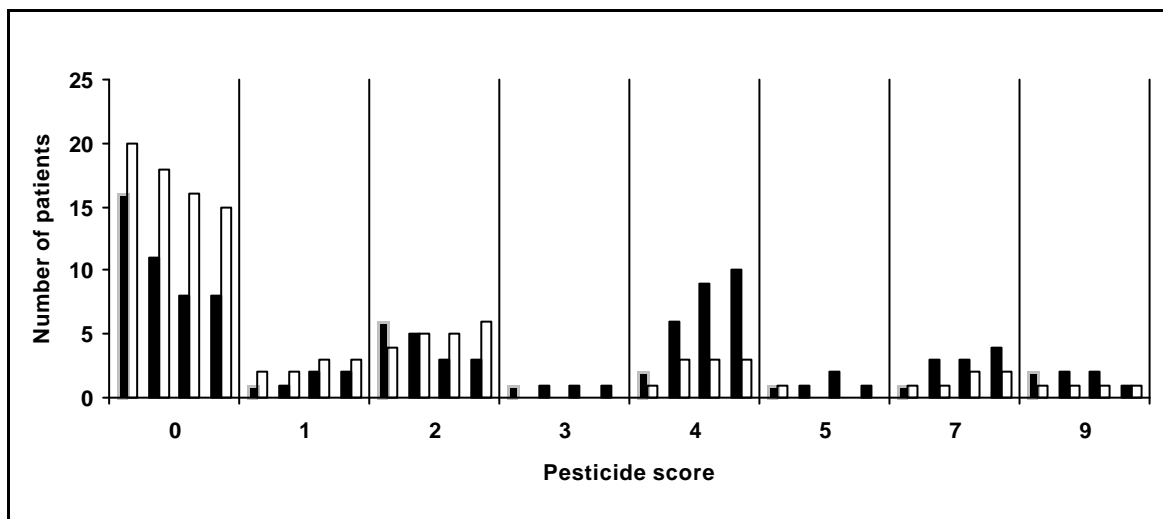


Figure 2: Histogram for pesticide exposure level, all periods successively described within each column, black boxes for cases, white boxes for controls.



4.7. X-rays

X-rays examinations are described using the following classification:

- lung (RX Thorax, scopy)
- stomach
- intestine (baryte)
- kidney
- angiography
- pelvis
- other bones
- tooth (one tooth or panoramic)

Using the following weights:

- X-rays of the Lung: 1
- From scopy to angiography: 50
- Rx pelvis: 50
- Other bone: 10
- Tooth:1
- Tooth panoramic: 20

and ignoring the other examinations, description of the X-rays examinations performed for cases and controls patients for each decade is provided in table 13.

Table 13: Mean X-rays code for case and control group.

	Period4	Period3	Period2	Period1
Cases	66.48	226.55 (57.59)*	70.34	65.38
Controls	52.07	58.87	94.40	69.97
P-value for case vs. control comparison	0.590	0.7947	0.5135	0.6814

* Within parentheses the observed mean ignoring patient 1 who had 108 kidney examinations in period 3 is also provided.

No statistically significant difference was detected for this global score, neither for the number of examinations performed within each category.

4.8. Dioxin exposure

The distance of work place and living environment from incinerator/hospital was recorded for each patient. For this analysis, within each patient, the minimal distance was taken into account, and dichotomised according to the 500m cut-off.

For period 4, among the 5 discordant pairs (i.e. the pairs for whom cases and controls exposure assessment differ), 5 controls (3 ulcers and 2 colons) were exposed, while their matched controls were not (p-value = 0.008). For period 3, among the 7 discordant pairs, 6 controls were exposed while their matched controls were not (p-value = 0.046). For period 2, 5 discordant pairs were observed, with 4 controls exposed (p-value = 0.16), and for period 1, 4 discordant pairs were observed with 3 controls exposed.

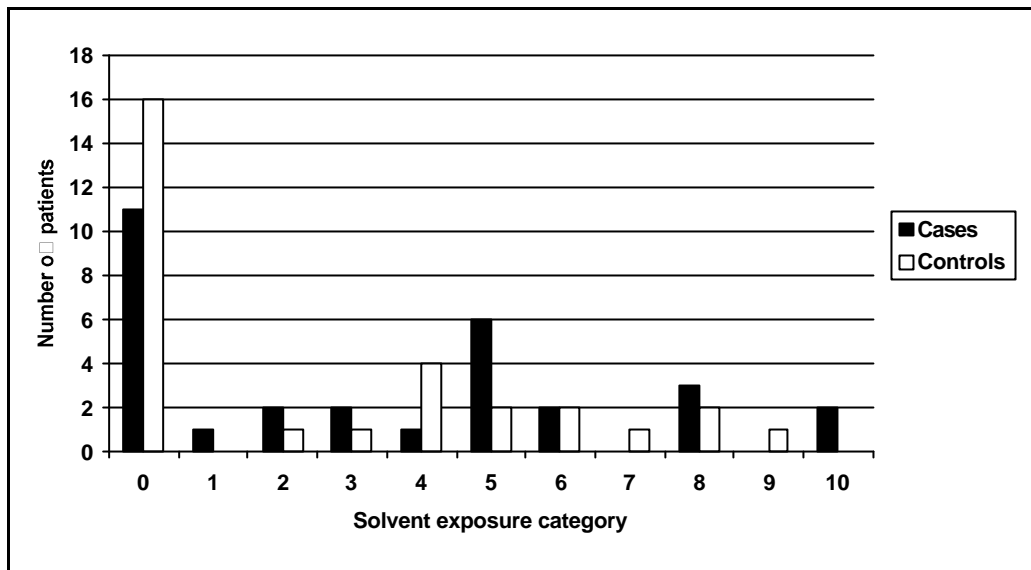
4.9. Solvent exposure

A score has been attributed to each subject by a panel of experts. The minimal score, 0 corresponds to very low exposure level while the maximal score corresponds to very high exposure level, as derived from the whole exposure history from each patient. Patients distribution for this score and according to their case/control status is given in table 14 and figure 3. No statistically significant effect was detected (p=0.256). Excluding the pairs corresponding to a matched kidney control, the p-value drops to 0.195.

Table 14: Patients distribution across solvent score category.

Score	0	1	2	3	4	5	6	7	8	9	10
Cases	11	1	2	2	1	6	2	0	3	0	2
Controls	16	0	1	1	4	2	2	1	2	1	0

Figure 3: Patients distribution within solvent score category.



4.10. Link with karyotypic information.

The karyotype information was available for 28 case patients, normal for 13(46%) of them, and abnormal for 15 (54%) of them, with the anomaly involving chromosome 5,7 or 11 in 8 of them. The 8 patients with an anomaly involving either of chromosomes 5, 7, 11 known to be associated with MDS have been studied separately and a gradient of exposure levels was observed for both the pesticide exposure score, and the drug consumption score, within each decade, but not for the solvent exposure score. This finding is described in table 15. In order to study this trend further, a model explaining the variability of the pesticide exposure level scores through the karyotype group (1 for control, 2 for cases without 5,7,11 anomaly, 3 for cases with 5,7, or 11 anomaly) was assessed and gave non-significant p-values of 0.263 and 0.069 respectively in periods 4 and 3, but significant p-values of 0.021 for period 2 and 0.018 for period 1. The corresponding relative risk point estimates and associated 95% confidence intervals are 3.04 [1.21,7.64] for period 2, and 3.00 [1.24;7.24]. A similar model for the medication did not prove statistically significant (p-value = 0.33).

Table 15: Mean score level for pesticide exposure and drug consumption, within control patients, case patients with no anomaly of the chromosomes 5,7, or 11 and patients with such anomalies.

	period 4	period 3	period 2	period 1
Pesticide code				
Control	1.17	1.33	1.60	1.67
cases with no anomaly of chromosome 5, 7, nor 11	1.15	2.65	3.00	2.75
cases with an anomaly in chromosomes 5, 7, 11	2.5	2.75	3.63	3.89
		All periods		
Medication code				
control		2.77		
cases with no anomaly of chromosome 5, 7, nor 11		3.05		
cases with an anomaly in chromosomes 5, 7, 11		3.38		
		All periods		
Solvent code				
control		2.50		
cases with no anomaly of chromosome 5, 7, nor 11		3.80		
cases with anomaly in chromosomes 5,7 or 11		2.87		

Annex: Medication list.

Some medication appear twice in the list due to technical problems during the data abstracting phase, but have correctly been taken into account during the analysis. These are dioxin (either 28,56), and chloramphenicol (49,50).

<i>Medication code</i>	<i>Includes</i>
1	GI Ulcus treatment
2	Allopurinol (Zyloric)
3	Antibiotics
4	Antiarythmica
5	Sedativa
6	Acetylsalicylic acid
7	Vitamin C
8	Vitamin D
9	Calcium
10	Sulfasalazyne
11	NSAIF(Non steroidal anti inflammatory farmaca)
12	Oestrogen
13	Progestogen
14	Paracetamol
15	Vasodilatantia
16	Anti-malaria
17	Anti-hypertensiva
18	Methylene blue
19	Anti-reuma
20	Indometacine
21	Insuline
22	Glaucoma treatment
23	Cortisone
24	Asthma treatment
25	Raloxifen
26	Caffein
27	Anticoagulantia
28	Digoxine
29	Bronchospasmoditica
30	Radiotherapy
31	Diuretics
32	Tanderil
33	Tuberculosis
34	Spasmoditica
35	Hyperlipemia treatment
36	Anti-anginosa
37	Antidiabetica other than insuline
38	Anti-migraine
39	Anti-emetica
40	Anticonceptiva
41	Anti-depressiva
42	Analgetica
43	Diverticulitis treatment
44	Myorelaxantia
45	Antimetabolites
46	Zantac
47	Vitamin B
48	Folum acid
49	Chloramphenicol
50	Chloramphenicol
51	Antitussiva
52	Anti-hemorragica
53	Anti-adipositas

54	Mucolytica
55	Alkylating agents
56	Digoxine
57	Paracetamol
58	Aluminium
59	Colchicine
60	Cerebral circulation treatment
61	Cholinesterase inhibitors
62	LHRH
63	anti-mycotica

Annex VIII

Study proposal

The relationship between genetic susceptibility factors, the peripheral blood cell count of workers exposed to an industrial myelotoxin, and the risk of MDS/leukemia

Investigators / Authors:

Dr Karel Van Damme

Dr. Ludwine Casteleyn

Dr. Alix Collard

1. Introduction

Preventing occupational diseases possibly due to benzene exposure, especially hematolymphopoietic disorders like MDS or leukemia are a major concern in many industries. The mechanisms of leukemia and MDS causation are not yet fully understood. Understanding the mechanisms of leukemia and MDS causation is important not only with a view to orienting preventive actions both in industry and living environment and to improving the surveillance of exposed workers, but even in order to develop better therapeutic procedures in the near future. To reveal these mechanisms, a series of studies have to be performed, each of them contributing to solving part of a possibly very complex puzzle. Despite dramatical improvement of working conditions in many of those industries, there is no guarantee as to the risk of severe adverse health effects having completely disappeared. This is why peripheral blood cell counting still is and has to remain a common test as part of periodical medical follow-up procedures for benzene exposed workers. But there is still uncertainty about the meaning of some possibly benzene related alterations in peripheral blood cell count as indicators of an increased risk for more severe effects.

Susceptibility factors are believed to play a role in disease causation and in therapeutic response, as well as in benzene-related hematolymphopoietic disorders and are the interest centre for this study.

Since it is well known that severe peripheral blood cell count disturbances, bone marrow dysplasia and some types of leukemia may all be related to benzene exposure, and because, for historical reasons, benzene is the best studied exposure factor in relation to these diseases (although not necessarily the most important), a benzene exposed population is a resource worth exploiting for contributing to answering in part some of the the questions we are interested in:

- do a series of particular known polymorphic genes play a role in susceptibility to some types of leukemia; to bone marrow dysplasia; to pathological changes in peripheral blood cell count; and to alterations in peripheral blood cell counts of healthy individuals who were benzene exposed ?
- Is the distribution of types of polymorphisms the same for all these possible effects? In other words, are these different effects probably related or not?
- How to interpret alterations in blood cell counts performed as part of health surveillance of exposed workers with respect to the risk of severe hematolymphopoietic disorders?

Individuals may react in very different ways to similar exposure situations and for different effects. This is obviously due to both intra- and inter-individual differences in susceptibility to particular effects. Susceptibility is the result of a complex interaction of inherited and acquired susceptibility affecting factors. Susceptibility to a particular effect is constantly evolving within any single individual. In order to study the possible influence of one of the factors involved in a complex metabolic process resulting

in effects, the theoretical ideal would be to have two groups with identical exposure and all susceptibility factors being equal, except the factor under study. The animal experiment is the ultimate paradigm for this type of research. This theoretical ideal is extremely far from what is feasible with human data. We do not consider here the research possibilities with experiments on humans, but the use of human data in epidemiological research. Epidemiological research aiming at studying exposure-response relationships requires the development of a methodology which allows to distillate such relationships out of an apparently chaotic complexity of interfering variables –only some of which are known- through a complex procedure of rearrangement and classification of data. This complexity not only resides in the fact that factors affecting metabolism are largely unknown or not sufficiently understood, but also in the difficulty to assess correctly exposure histories, and even more so to assess the resulting dose which often cannot be correlated in a reliable way with exposure. These are nevertheless the uncertainties we have to deal with when studying effects..which on their turn are not completely understood.

The general objective of this study proposal is to generate hypotheses on the relationship between reversible changes in peripheral blood cell counts in benzene exposed workers and the risk of severe hematolymphopoietic disorders, and to learn more about presumed genetic susceptibility factors possibly affecting these effects. Two main populations will be studied consisting of benzene exposed workers on the one hand and of patients suffering from myelodysplasia/leukaemia on the other

2. Study material

2.1. Benzene exposed workers

The study base consists of all employees and retirees of a petrochemical plant who had possible exposure to benzene and for whom at least 5 peripheral blood cell counts were collected before the end of 1989. Since, in the petrochemical plant considered, there has been neither a tradition of exclusion of workers for reasons related to the material of interest to this study, nor a tradition of selecting workers for genetic polymorphism, or excluding them for any possible health problem related to benzene exposure, this population can be considered as unselected with respect to the genetic polymorphisms under study. The study population will be selected from the study base as explained in the section 'methods'. The study base and database on peripheral blood samples is a result of a study on "Cancer incidence in chemical and petrochemical workers" (Impulse programme 'health risks' Contract NR HH/06/038).

The following information is available:

- ❑ Blood cell counts: a database with 13.539 peripheral blood cell counts of 360 benzene exposed workers was selected from the individual medical records archived by the occupational health physician of a petrochemical plant. The data of origin included 17.404 peripheral blood cell counts taken from 1966 till 1989 of 457 potentially benzene exposed workers. Only persons who were first employed before 1978 were included in the selected database. Persons who contributed less than 5 blood samples were not included. Also control blood samples following cell counts out

of normal ranges were not included. For each individual included, the following data are available: code number, date of birth, total number of samples, and results for each sample, including: date of sample taking, months since first sample, interval with previous sample in months, haemoglobin, hematocrit, RBC (red blood cell count), platelets, WBC (white blood cell count), and in most cases WBC differentiation in neutrophils, granulocytes, lymphocytes;

- ❑ Benzene exposure data: general exposure data are derived from stationary air samples (n=3867), collected from 1978 onwards;
- ❑ Other possibly relevant individual medical records as kept by the OHP (occupational health physician) for instance on drug use, data on individual smoking habits, and on job category in different time periods.

Since no information is available in this database about genetic polymorphisms, results of tests for genetic polymorphisms will be collected from those selected members of the study population willing to participate according to the procedure explained in section 'ethical considerations' The analysis of the samples will be done at the Finnish Institute of Occupational Health (laboratory of Dr A Hirvonen). At least the following polymorphic genes will be considered in the study:

- GSTM1 and GSTT1 coding for the glutathione S-transferases M1 and T1, a family of isoenzymes catalysing the conjugation of reactive species to glutathione. Individuals with the GSTT1 null genotype were suggested to have enhanced susceptibility to MDS (Chen et al, 1996);

- CYP 2E1 coding for P450 2E1, an enzyme involved in the metabolism of several compounds including benzene.

- NQO1 coding for the NQO1 enzyme converting the benzoquinones, which are potent hematotoxic and genotoxic benzene metabolites, back to their less toxic hydroxy metabolites. A higher frequency of NQO1 inactivating polymorphisms has been identified in subjects developing hematotoxicity after benzene exposure.

2.2. Patients suffering from myelodysplasia/leukaemia

The study base consists of all adult patients suffering from myelodysplasia or a type of benzene-related leukaemia which were diagnosed and/or treated in the haematology service of one of the hospitals which are part of the network set up during the study on 'Exposure specific genomic lesions in myelodysplastic syndromes'- (Programme of scientific support to worker protection in the area of health, contract nr. ST/03/23). These patients will be invited to participate. Those who accept are the

study population. Extending this network and thus the number of patients who could possibly be included is an aim of this study proposal.

All patients who agree to participate, will give a blood sample of 2 ml for the analysis of the same genetic polymorphisms as carried out for benzene exposed workers.

3. Methods

3.1. Categorising the benzene exposed workers

In studying the relationship between genetic polymorphisms and susceptibility to benzene induced effects, maximising the reliability of the classification of study base members in effect categories as a function of exposure history is both crucial and complex. Limiting the number of study base members included in these categories, thereby increasing the contrast between effect categories, will be done in order to increase the sensitivity of the study. A procedure for selecting individuals from the population is being developed based on two complementary approaches: a subjective interpretation of data by experts on the one hand (1), and a mathematical approach (2) on the other. The process for developing this procedure consists of comparing both approaches with a view to improving mutual consistency of both.

Subjective expert interpretation.

For every individual included in the study base the cell counts history will be interpreted by medical researchers who are trained in occupational health and by the occupational health physician who collected the data and keeps the individual medical records, in order to distinguish different types of changes in blood cell count history, as a function of assumed individual benzene exposure history. With respect to exposure history, approximately, the following discontinuous periods of benzene exposure will be considered:

- the late sixties and early seventies, when the plant was started up with relatively high benzene exposure. In the same period, benzene was frequently used to wash hands, with skin absorption as a probable consequence (period 1);
- the mid and late seventies, when benzene exposure was lower than during the early phase, and washing hands in benzene was less common (period 2);
- the mid eighties¹, when benzene exposure was falling rapidly to reach very low levels as a time weighted average and washing hands in benzene was not done (period 3).

In order to classify individuals, this general assessment of exposure to benzene in different time periods of observation will be completed further by individual exposure history assessment, starting

from considering historical air sampling data, department and activity altogether, and other possibly relevant information on the individual the occupational health physician is aware of such as disease histories of individuals and subsequent medication. The latter is important because it may be a possible confounder in studying the relationship between blood cell counts and exposure to benzene.

In view of these exposure assessments, individuals will in a first exercise be categorised with respect to their cell count histories and in comparison with a reference population as follows:

- individuals who consistently showed low white blood cell counts in period 1 and/or 2, and higher near average values in period 3;
- individuals who consistently showed high white blood cell counts in periods 1 and/or 2, and lower near average values in period 3;
- individuals who hardly showed any variability in WBC counts in periods 1, 2 and 3;
- individuals who showed a high variability in WBC counts with no apparent relationship with exposure periods;
- individuals who rather consistently show low lymphocyte counts in period 1 and/or 2, and higher near average values in period 3;
- individuals who consistently show high lymphocyte counts in periods 1 and/or 2, and lower near average values in period 3;
- individuals who showed cell count changes for other than white blood cells, and which are more pronounced in the earlier periods;

Mathematical approach.

The results of this classification into effect categories carried out by medical and occupational health experts is then compared with a mathematical approach using the same data. A trial-and-error process is used to develop the approach, starting from a mathematical translation of the initial formulation of the category, and continuing with a comparison between the workers that should have been classified in the category, and the workers actually classified in it, in order to understand the reason for the discrepancies and to correct the criteria for them. The process is stopped as soon as the possible discrepancies between the target list established by the medical experts and the list established using the defined criteria is considered negligible by the medical professionals.

3.2. Analysis of the data on benzene exposed workers

In order to assess the frequency of genetic polymorphisms in a reference population, sampling will be done amongst all current employees who gave a blood sample for genetic analysis. The sampling size will be in accordance with the expected frequency of occurrence of each polymorphism under study. Genetic polymorphisms possibly related to susceptibility will be compared between the categories, by means of a Fisher's exact tests for the hypothesis that each particular category of

benzene effects as defined by the health professional experts might be associated with a particular (combination of) polymorphism(s). Relationships between peripheral blood cell count, benzene exposure and particular genetic polymorphism's will thus be studied, starting with the above described approach. Such an approach was considered the most appropriate with a view to generating hypotheses. Hypothesis testing may be done subsequently. The complexity of the problem and the natural variability or uncertainty of the data being dealt with require a strategy which enables clear hypotheses to be generated before choosing the statistical model to deal correctly with the complexity of the data and thus to test this hypothesis. This cascade method of proceeding enables to limit the number of persons to be tested for genetic susceptibility genes at each step of the decisional cascade. Hypotheses resulting from this procedure will be tested on larger groups of well-selected persons in the database. The whole procedure will be called a "cascade exploratory approach".

The whole research team will reconsider all findings and then decide on the next steps. These may include extension of the study population and the blood samples data base. In this way, genetic polymorphism's may be identified which can be considered as relevant for susceptibility to particular types of reversible effects on peripheral blood cell count.

3.3. Comparison with the polymorphisms found in patients suffering from myelodysplasia/leukaemia

The frequency of occurrence of these polymorphisms, alone and in combination, in the different categories of benzene exposed workers, will be compared with their frequency in a population of persons suffering from myelodysplasia/leukaemia, using both a classical and a Bayesian statistical approach for proportions comparison. Subgroups of myelodysplasia/leukaemia patients may be used in the analysis. The possible relationship between particular types of changes in peripheral blood cell count and the risk of severe hematolymphopoietic disorders may be assessed subsequently.

4. Ethical considerations:

Workers:

Democratic participation:

The research team has already obtained the unanimous agreement of workers and employers representatives in the Committee for Prevention and Protection (previous health and safety committee) and the full collaboration of the occupational health service to participate in this study.

Informed consent:

The procedure for collecting samples for analysis is as follows: all individuals coming to periodical medical examination will be invited by the occupational health physician to give a blood sample of 2 ml for the analysis of genetic polymorphisms. The individuals will be informed both orally and in written form about the aim of the study. The individuals willing to participate will sign an agreement to this effect. The content of the agreement form has been discussed with the parties involved and approved.

Confidentiality:

The samples will immediately be marked with a code number only. Only the occupational health physician and two members of the research team (KVD and LC) have access to the code. Coded samples will be stored by the occupational health physician until transportation to the Finnish Institute of Occupational Health (laboratory of Dr A Hirvonen), where the analysis will be performed. The occupational health physician will participate in the analysis, and may know about the individual genetic test results. This procedure has been fully approved by the same Committee.

Information:

The Committee for Prevention and Protection will be informed of the results of the study.

Patients:

Patients will be informed both orally and in written form about the aim of the study. The treating physician either in the hospital, or the general practitioner- will invite the patient to participate. The general practitioner will be informed, or his agreement will be requested, in accordance with existing protocols of the ethical committees. Individuals who are willing to participate in the study will be asked to sign an agreement to this effect. The whole procedure will be subjected to the ethical committees of the hospitals involved. Samples will be marked with a code number only. They will be stored at the haematology service until transportation to Finland.

Two researchers, (KVD and LC) have access to the code. The treating physicians have access to the individual results of their patients if the ethical committees and the patients do agree.