



Inventarisatie van Koppelwensen en mogelijkheden van Biobanken en Databanken in Nederland

Auteurs: Mark C.H. de Groot¹, Willem de Bruijn^{1,2}, Gerard van Grootheest^{3,4}, Olaf H. Klungel¹,
Jan H. Smit^{3,4}.

1 Universiteit Utrecht, Utrecht

2 Stichting Mondriaan Health Research Data, Utrecht

3 GGZ inGeest, Amsterdam

4 VU Medisch Centrum, Amsterdam

19 maart 2014

INHOUD

Inleiding	3
Methode	3
Resultaten	4
Samenvatting en conclusies	11
Referenties	13
Bijlage	13

INLEIDING

Het Biolink NL project (BBMRI-NL Regenboog 5 project) heeft tot doel de doelmatigheid en validiteit van koppelingen tussen biobanken en registers te verbeteren. Onderdeel hiervan is het inventariseren van de wensen en mogelijkheden van biobanken en registraties met betrekking tot het koppelen aan andere gegevens. Deze inventarisatie omvat de technische mogelijkheden tot koppeling, juridische voorwaarden, beperkingen en waarborgen, informed consent, toegevoegde waarde en wensen van koppeling. Om te voorkomen dat beheerders van biobanken en registers onnodig worden belast met lange vragenlijsten is de vragenlijst voor inventarisatie van de koppelingen strikt beperkt tot dat domein; achtergrondinformatie betreffende de registers is beschikbaar via BBMRI-NL en BBMRI-EU vragenlijsten die in het verleden zijn verkregen.

METHODE

Vragenlijst

De vragenlijst is ontwikkeld volgens bovenstaande uitgangspunten en geredigeerd in een aantal ronden waarbij interne en externe ervaringsdeskundigen rondom registraties zijn geraadpleegd. Gekozen is voor een digitaal Adobe PDF formulier wat eenvoudig offline elektronisch in te vullen was en waarvan de antwoorden geautomatiseerd verwerkt konden worden. De respondenten hadden ook de mogelijkheid om het formulier af te drukken en handmatig in te vullen (zie bijlage voor de volledige vragenlijst).

Adressering

Voor het verzenden van de vragenlijst is gebruik gemaakt van de adreslijst van BBMRI-NL waarin alle bij BBMRI-NL geregistreerde biobanken vermeld staan.

Na opschonen en ontdubbelen konden 152 biobanken en 36 registraties aangeschreven worden waarvan er een emailadres achterhaald werd via telefonische navraag of internet zoekacties.

RESULTATEN

Respons aantallen

De vragenlijst is op 10 januari 2013 verstuurd aan 152 BBMRI-NL biobanken en 36 belangrijke, niet aan BBMRI-NL verbonden, registraties.

Tabel 1 Ontvangen respons na verzending en herinneringen

Bron	Verstuurd	Initiele respons	Na email herinnering	Na telefonische herinnering	Geen reactie	Bruto respons %
Biobanken	152	34	50	53	78	34,9
Databanken	36	9	15	15	16	41,7
Totaal	188	43		68	94	36,2

De reacties op deze vragenlijst waren overwegend positief en een deel van de respondenten heeft aangegeven graag op de hoogte gehouden te willen worden van de resultaten. In een aantal gevallen is bekend geworden waarom er geen ingevulde vragenlijst retour is ontvangen. Deze informatie valt te categoriseren zoals in onderstaande tabel. Een aantal biobanken waren incorrect geregistreerd in het adresbestand, bestaan niet meer, zijn het land uit verhuisd of waren uiteindelijk niet te traceren tot een verantwoordelijk persoon. Hieronder vallen bijvoorbeeld ook verzamelingen die geen betrekking hadden op personen, zoals een verzameling bacteriestammen of personen die niet in Nederland woonachtig zijn. Een zestal registraties gaf aan de vragenlijst niet in te kunnen vullen omdat zij zich in een opstartfase bevinden of aangaven geen identificerende gegevens te hebben of technische mogelijkheden tot koppeling te zien. In een aantal gevallen (13) was er sprake van duplicaat of overlap waarbij deelverzamelingen van subjecten binnen een biobank oorspronkelijk als verschillende registers of onder verschillende naam, waren aangemerkt. Als we bovengenoemde administratieve onvolkomenheden buiten beschouwing laten kunnen we stellen dat van 162 “terecht verstuurd” enquêtes er 68 (42%) zijn geretourneerd als netto respons.

Tabel 2 Analyse Niet-responders

Bron	Verstuurd	Niet meer bestaand	Duplicaat / overlap	Kan (nog) niet	Terecht verstuurd	Ingevuld	Netto respons %
Biobanken	152	5	13	3	131	53	40,5
Databanken	36	2	0	3	31	15	48,4
Totaal	188	7	13	6	162	68	42

Antwoorden vragenlijst

Delen van informatie

Om registraties te kunnen verrijken voor wetenschappelijk onderzoek door te koppelen met andere bronnen is het van belang om inzicht te hebben in de attitude van de beheerder/eigenaar van de te koppelen bron met betrekking tot koppelbereidheid. Van belang kan hierbij zijn wat het oorspronkelijke doel van de gegevensverzameling is.

Tabel 3. Met welk primair doel worden of zijn de gegevens of lichaamsmaterialen verzameld?

Doel gegevensverzameling	Aantal	%
Onderzoek	59	86,8
Diagnostiek-Onderzoek	5	7,4
Behandeling	1	1,5
Declaratie, Statistieken	2	2,9
Missing	1	1,5

De bereidheid om informatie met derden te delen voor wetenschappelijk onderzoek is hoog onder zowel biobanken als databanken, er is geen registratie die hier bezwaar tegen maakte.

Tabel 4. Bestaat de bereidheid om biomaterialen of data onder voorwaarden te delen met derden voor wetenschappelijk onderzoek en verschilt dit voor databanken en biobanken?

	Aantal	% Biobank	% Databank
Ja	54	79,6	80,0
Ja, onder voorwaarden	13	18,5	20,1
Nee	0	0,0	0,0
Missing	1	1,9	0,0

Op de open vragen of er een beoordelingsprocedure of protocol bestaat en welk bestuurlijk orgaan of persoon dergelijke aanvragen voor gegevens of materiaal beoordeelt, werd verschillend en overlappend geantwoord. Samenvattend laten de antwoorden zien dat er in 72% van de registers iets geregeld en vastgelegd is in een procedure.

Tabel 5 Samenvatting vragen naar protocol, procedure of instantie voor delen informatie en materiaal.

	Aantal	%
Ja, protocol	37	54,4
Ja, aanvraag indienen bij instantie	11	16,2
Nee, er is niets vastgelegd	19	27,9
Missing	1	1,5

Bereidheid tot koppelen van informatie

Vragen over het koppelen met andere bronnen zijn uitgesplitst naar bereidheid om te koppelen, bestaande koppelingen en met wat voor type gegevens er reeds gekoppeld wordt en wat de koppelwensen voor de toekomst zijn.

De bereidheid tot koppelen is hoog. Als de antwoorden “nee” en “kan niet” worden opgeteld geeft slechts 10,3% aan nu niet te koppelen of dit misschien te willen gaan doen.

Tabel 6 Bereidheid om te koppelen, reeds bestaand en in de toekomst

	Aantal	%
Ja	57	83,8
Nee	4	5,9
Misschien in toekomst	3	4,4
Kan niet volgens respondent	3	4,4
Missing	1	1,5

Bestaande koppelingen die in de open vragen genoemd werden zijn in 54,4% van de gevallen koppelingen met ook externe, in de betekenis van, buiten (de database van) de organisatie,

verzamelde gegevens. Er is niet specifiek gevraagd waar, dat wil zeggen op welke locatie, die koppeling plaatsvindt.

Tabel 7 Bestaande koppelingen samengevat naar soort

	Aantal	%
Ja (niet gespecificeerd)	1	1,5
Ja, intern	11	16,2
Ja, intern en extern	37	54,4
Geen bestaande koppelingen	18	26,5
Missing	1	1,5

Een analyse van de antwoorden op de open vragen waarmee men dan zou willen gaan koppelen laat zien dat bepaalde soorten informatie zeer in trek zijn. Medicatiegebruik en/of klinische bestanden zijn samen het vaakst genoemd als wens in meer dan 90% van de antwoorden. Het CBS werd vaak als entiteit genoemd, zonder specificatie van het gewenste CBS register.

Tabel 8 koppelwensen naar type informatie

	Aantal	%
CBS	1	1,5
Medicatie records	7	10,3
Klinische records	5	7,4
Medicatie+Klinische data	4	5,9
CBS+Medicatie+Klinische data	17	25,0
Lab, DNA, Med Test	2	2,9
Kanker registratie	4	5,9
Geen of niet specifiek	24	35,3

Beschikbare koppelvariabelen

Om te koppelen wordt gebruik gemaakt van koppelvariabelen. In de inventarisatie hebben we gevraagd hoe deze verkregen en vastgelegd worden. De nauwkeurigheid waarmee deze

variabelen worden vastgelegd bepaalt in hoge mate het succes waarmee gekoppeld kan worden. 1,2 In 63% is er sprake van enige vorm van validatie van deze koppelvariabelen, bijvoorbeeld controle door herhaalmetingen, zorgverlener of controle met de gemeentelijke basisadministratie.

Er is geïnventariseerd welke identificerende variabelen er in de diverse registraties aanwezig zijn. In tabel 9 zijn de meest relevante koppelvariabelen en de beschikbaarheid in de registraties vermeld.

De aanwezigheid van een adreshistorie is van belang bij het koppelen van personen waarbij het tijdstip van dataverzameling van belang is zoals bij longitudinale of retrospectieve onderzoeken.

Tabel 9 Selectie van enkele beschikbare koppelvariabelen.

	Aantal	%
Burger Service Nummer	15	22,1
Geboortejaar	60	88,2
Geboortedatum	54	79,4
Geslacht	62	91,2
Achternaam	39	57,4
Postcode (4 tekens)	31	45,6
Postcode (6 tekens)	31	45,6
Adres	33	48,5
Adres historie	16	23,5

Om te koppelen met externe registraties wordt gebruik gemaakt van een koppelsleutel die bestaat uit de combinatie van koppelvariabelen. De volledige in Biolink WP3 gebruikte koppelsleutel bevattende geboortedatum, geslacht, achternaam en postcode 4 is in 29,4% van de registraties aanwezig. De combinatie geslacht, achternaam en geboortedatum is in 51,5% van de registraties aanwezig. Indien specifiek gekeken wordt naar de beschikbaarheid van bovengenoemde combinaties in registraties die aangegeven hebben dat ze koppelen of in de toekomst willen gaan koppelen, dan vallen deze percentages hoger uit (tabel 10). Van de 60 registraties met koppelwens beschikt 22% niet over ofwel BSN ofwel een van de andere combinaties van koppelvariabelen.

Tabel 10 Beschikbaarheid van combinaties van koppelvariabelen gerelateerd aan alle registraties en aan registraties met koppelwens.

	Aantal	% alle	% koppelwens
Combinatie geslacht + achternaam + postcode 4 + geboortedatum	20	29,4	33,3
Combinatie achternaam + postcode 4 + geboortedatum	20	29,4	33,3
Combinatie geslacht + achternaam + geboortedatum	35	51,5	58,3
Combinatie geslacht + achternaam + postcode 4	21	30,9	35,0
Combinatie geslacht + postcode 4 + geboortedatum	28	41,2	46,7
Geen BSN, geen van bovenstaande combinaties koppelvariabelen	13	19,1	21,7

Beveiliging data en informed consent in relatie tot koppeling

De diverse maatregelen die genomen zijn om de privacy te beschermen staan weergegeven in onderstaande tabel. Op geaggregeerd niveau blijkt dat 89,7% minimaal één maatregel neemt, 45,6% minimaal een juridische en 82,4% minimaal technische maatregel zoals pseudoniemen of anoniem gebruikt.

Tabel 11 Gebruikte maatregelen om privacy van personen te beschermen bij het delen van data

Doel gegevensverzameling	Aantal	%
(Juridisch) Privacyreglement	24	35,3
(juridisch) Aanmelding CPB	19	28,4
(technisch) Pseudoniem-zonder TTP	24	35,3
(technisch) Pseudoniem-met TTP	19	28,4
(technisch) Anonimiseren	37	54,4
Andere maatregelen	10	14,9

Informed consent is in 61,8% van de registraties gebruikt om iets over koppelingen vast te leggen. Bij 20,9% is er sprake van een opt-out regeling en totaal is er in 91,2% iets vastgelegd over het mogen koppelen. Uit reacties in de open vragen blijkt wel dat vaak onduidelijk is wat precies wel en niet mag binnen de afspraken.

Benodigdheden voor verbetering gebruik data en koppeling

Naast de vragen naar koppelingswensen hebben we aan het eind de volgende open vraag gesteld: 'Naast de mogelijk eerder door u genoemde wensen voor koppeling kan het zijn dat andere zaken nodig zijn om het (her)gebruik en koppeling van de data te faciliteren. Wat is hiervoor nodig naar uw mening binnen uw organisatie of op nationaal niveau?'

In de 33 antwoorden zijn 46 benodigdheden genoemd die als volgt zijn te categoriseren:

Tabel 12 Welke zaken binnen organisatie of landelijk niveau zijn nodig om hergebruik en koppeling van de data te faciliteren?

	Aantal	%
Specifieker of ruimer Informed Consent / aanpassing regelgeving	5	10,9
Universele Trusted Third Party	9	19,6
Universeel datamodel voor uitwisseling/ datamanagement	5	10,9
Organisatorische faciliteiten / Governance	7	15,2
Catalogus met beschikbare registraties en variabelen	4	8,7
Juridisch kader wat duidelijk maakt wat wel / niet mag	4	8,7
BSN via TTP of versleuteld BSN toestaan en beschikbaar maken	10	21,7
Geld	2	4,3

Opmerkelijk is hier dat de grootste behoefte bestaat aan technische verbeteringen in de vorm van koppelmogelijkheden via TTP, koppelvariabelen en/of BSN en de bijbehorende kennis en dat financiële middelen slechts 2 maal genoemd werden.

SAMENVATTING EN CONCLUSIES

In deze inventarisatie is gekeken naar de koppelbereidheid, koppelwensen en koppelmogelijkheden van een groot aantal Nederlandse biobanken en databanken. Opmerkelijk is de zeer hoge bereidheid om materiaal en data te delen voor wetenschappelijk onderzoek en 72% heeft daar procedures of protocollen voor. Koppelbereidheid is ongeveer 90% en de grootste interesse bestaat er voor klinische en medicatie gegevens.

De technische mogelijkheden verschillen per registratie als het gaat om koppelvariabelen, naam en adresgegevens zijn niet universeel aanwezig. De borging van de privacy bij koppeling en delen van informatie heeft de aandacht van de registraties.

Deze inventarisatie is gehouden onder een selecte groep, namelijk biobanken aangesloten bij BBMRI en databanken met een onderzoek reputatie. De respons op de verstuurde enquêtes is 40%. Dit betekent dat de mogelijkheid bestaat dat er een overmatige representatie bestaat in de antwoorden van registraties die welwillend staan tegenover delen en koppelen. De antwoorden maken het wel mogelijk om selecties te maken voor koppelprojecten.

Toekomstige inspanningen om koppelingen te faciliteren

Uit deze inventarisatie blijkt dat er een hoge bereidheid is om informatie te delen en te koppelen met andere registraties en dat dit al op ruime schaal gebeurt. De geconstateerde beperkingen liggen op het gebied van de koppelvariabelen en de juridische mogelijkheden c.q. onduidelijkheid om te kunnen en mogen koppelen en staan samengevat in tabel 13.

Registraties met bestaande koppelingen zijn over het algemeen ook de registraties met protocollen voor delen en hebben de meeste voorzieningen getroffen rondom vastlegging van toestemming voor koppelen. Vaak genoemd als verbetering voor koppelen zijn technische voorwaarden zoals overzicht (catalogus) en beschikking van de juiste koppelvariabelen en daaraan verbonden TTP en conditioneel gebruik van BSN. Verder bestaat er behoefte aan juridisch kader en governance, het is soms veel gevraagd voor individuele registraties om dit allemaal zelf te organiseren.

Toekomstige inspanningen voor het verbeteren van (her-)gebruik van beschikbare informatie in registraties door deze via koppeling te verrijken, zullen zich dan ook hierop moeten richten. BBMRI-NL kan met Biolink-NL zorgen voor kennis over het juridisch kader en hoe te koppelen.

Registraties zullen zelf aan de slag moeten om hun informed consent en toestemmingen te regelen en de juiste koppelvariabelen vast te leggen.

Tabel 13 Status van koppelwensen en mogelijkheden en aandachtspunten

	Bereidheid delen en koppelen	Koppelvariabelen	Juridisch/privacy
Aanwezig	Hoge bereidheid tot delen biomaterialen en gegevens en tot koppelen	Nagenoeg alle registraties hebben tenminste één combinatie van koppelvariabelen beschikbaar die geschikt is voor de meeste vormen van koppeling.	Nagenoeg alle registraties hebben maatregelen genomen om de privacy te beschermen.
Aandachtspunt	Veel registraties hebben iets van een procedure of systeem voor de beoordeling van aanvragen. Een vaste beschreven procedure is nog niet voor alle registraties beschikbaar.	Niet altijd is er de beschikking over de juiste combinatie van koppelvariabelen om een goede koppelsleutel te maken.	Koppelen is lang niet altijd vastgelegd in het informed consent. Er is veel onduidelijkheid over de juridische kaders.

REFERENTIES

1. Newcombe HB, Fair ME, Lalonde P. Discriminating powers of partial agreements of names for linking personal records. Part I: The logical basis. *Methods of information in medicine* 1989;28(2):86-91.
2. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *Journal of clinical epidemiology* 2011;64(5):565-72.

BIJLAGE

Bij dit document hoort als bijlage de vragenlijst verstuurd aan biobanken en databanken.