For reprint orders, please contact:
reprints@futuremedicine.com

Healthy People 2010 disease prevalence in the Marshfield Clinic Personalized Medicine Research Project cohort: opportunities for public health genomic research

CA McCarty[1†], BN Mukesh[1], PF Giampietro[1] & RA Wilke[1]

[†]Author for correspondence
[1]Marshfield Clinic Research Foundation, 1000 N. Oak Avenue (ML1), Marshfield, WI 54449, USA
Tel.: +1 715 389 3120; Fax: +1 715 389 4950; E-mail: mccarty.catherine@mcrf.mfldclin.edu

Objectives: The purpose of this study was to estimate the prevalence of Healthy People 2010 disease conditions in a large population-based cohort in central Wisconsin (WI, USA) and to consider how these conditions can be prioritized for research based on the use of healthcare services, the prevalence of various disease states and the resulting study power. **Methods:** Healthy People 2010 diagnoses were estimated for participants in the Personalized Medicine Research Project (PMRP), a large population-based biobank for residents aged 18 years and older living in central Wisconsin. By interrogating the electronic medical record, three parameters were calculated for each diagnosis: mean number of concomitant diagnoses, mean number of annual clinic visits before diagnosis and mean number of clinic visits after diagnosis. **Results:** A total of 18,239 adults enrolled in PMRP from September 2002 to May 2005 and were included in the study. They had a mean age of 49 years (standard deviation: 18.5), ranging from 18–98 years; 57% were female. At least one Healthy People 2010 disease was diagnosed in 86.4% of the participants; 13.6% had never been diagnosed with any of these conditions. The median number of diagnoses per subject was three (range: 1–15). The median number of annual visits after diagnosis was lowest for chronic obstructive pulmonary disease (9.1) and highest for sleep apnea (17.9). Subjects with a diabetic retinopathy diagnosis had the highest number of concomitant diagnoses (mean: 6.8). **Discussion:** All of the diseases within the Healthy People 2010 list are purported to have at least some genetic component, with the exception of injuries. The PMRP cohort is large enough that diseases of public health importance can be studied in the context of a variety of clinical and environmental covariates. This database is being developed as a national resource and is particularly useful where the estimated disease prevalence is 5% or greater. For less common diseases, additional cases can be recruited from throughout the Marshfield Clinic system of care, with population-based controls selected from the main PMRP study cohort.

Public health genetics is defined as the application of advances in genetics and molecular biotechnology to improve public health and disease [1]. Genome-wide association studies, made possible through large collections of DNA samples, have the potential to improve diagnosis, treatment and prevention of disease through the identification of genetic variants [2]. In the USA, the Centers for Disease Control and Prevention have developed a strategic plan for translating advances in human genetics into public health action, and one of the goals is to assess how risk for disease and disability is influenced by the interaction of human genetic variation with modifiable risk factors [101]. One proposed solution has been the development of robust strategies to identify the genetic contributions to disease and drug response within large population-based cohort studies [3].

Many research groups have taken up the national and international challenge to improve community health through the creation of DNA biobanks to facilitate genomics research. The Marshfield Clinic Personalized Medicine Research Project (PMRP) is currently the largest population-based DNA biobank in the USA [4]. Established in 2002, the PMRP was created to facilitate research in the areas of pharmacogenetics, genetic epidemiology and population genetics, with the ultimate goal of improving population health and patient care. Nearly 18,000 adults participated during the first 24 months of subject enrollment. Resources and guidelines for tissue access are available on the PMRP website [102].

Although a study cohort of 18,000 is large, power may be limited for disease states occurring at a very low frequency. It is necessary to

Keywords: genomics, health priorities, healthy people programs, public health informatics

future medicine part of fsg

determine research priorities for this database and the Healthy People 2010 document represents one plausible starting point [103]. Although the Healthy People 2010 targets relate primarily to mortality and use of preventive screening services, these conditions have been identified as the most important in terms of overall public health priorities within the USA. All of the diseases within the Healthy People 2010 list are purported to have at least some genetic component, with the exception of injuries. In addition, many states, including the state of Wisconsin, have developed their own state-specific health priorities for the year 2010 [5]. Genetic predisposition to disease and family history have been ranked as the primary risk factors because of their relation to the 54 identified priority health conditions for Wisconsin.

The purpose of the current study was to estimate the prevalence of Healthy People 2010 conditions in a large study cohort in central Wisconsin and to consider how these conditions can be prioritized for genomic research based on prevalence, concomitant diagnoses and the use of healthcare services in a large, multispecialty group medical practice. Implications for future targeted recruitment have been considered in the context of power to detect associations.

Methods

The PMRP is a population-based cohort study with stored DNA, plasma and serum [4]. The study cohort ranged in age from 18–99 years at the time of consent, and is 57% female and 98% Caucasian; 76% reported German ancestry. They have been shown to be representative of the central Wisconsin adult population [4]. Participating subjects provided written informed consent to allow access to their Marshfield Clinic comprehensive electronic medical records. The Marshfield Clinic is an integrated regional healthcare system with 700 physicians in 41 locations throughout central and northern Wisconsin. All major medical specialties and subspecialties, except whole-organ transplant, are represented within the Marshfield Clinic system of care [104].

Detailed study methodology for the PMRP has been published previously [5]. Participant enrolment commenced on 18th September 2002. Initial recruitment was targeted to people aged 18 years and older who resided in one of the 19 zip codes around Marshfield (WI, USA), and for whom at least one member on their Marshfield Clinic account had received care at the Marshfield Clinic in the previous 3 years. The targeted 19-zip code area is known as the Marshfield Epidemiologic Study Area (MESA) [6]. Except for the city of Marshfield (population approximately 19,000 people), MESA residents reside rurally or in small towns or villages. The annual in- and out-migration is very low, making it ideal for prospective studies. By sharing an electronic medical record with neighboring hospitals, in-patient and out-patient diagnoses and procedures are captured for MESA residents. The Marshfield Clinic owns a health plan (Security Health Plan) to which many MESA residents belong, allowing capture of diagnostic codes for visits that occur outside the Marshfield Clinic system of care.

After providing written informed consent to participate in the PMRP, subjects completed a short baseline questionnaire to capture basic demographic information, smoking and alcohol intake and family history of common diseases and adverse drug reactions. The Research Project Assistant measured and recorded height and weight. Body mass index was subsequently calculated. Participants were defined as residents who signed informed consent documents and provided a blood sample for DNA extraction and storage and plasma and serum storage. The Marshfield Clinic Institutional Review Board approved all study procedures.

For the present study, all diagnostic and procedure codes contained in the combined electronic medical record of the Marshfield Clinic were extracted electronically. All diagnostic codes were extracted for the period from January 1 1960 to May 31 2005 to quantify the total number of visits. Only codes from the International Classification of Disease, 9th revision, (ICD-9) were used. Healthy People 2010 disease diagnoses were used for disease classifications [103]. Date of first diagnosis was identified, as well as the annual number of clinic visits.

'Rule of one' was used to classify subjects who received only one diagnosis of a given disease. To improve the specificity of the ICD-9 code electronic-based case finding, diagnoses made on two or more occasions ('rule of two') were considered as confirmation of that particular disease condition. Prior phenotype validation efforts have demonstrated that, for chronic conditions such as Type 2 diabetes, 'rule of two' is more accurate because many false-positive electronic diagnoses associated with visits to 'rule out' potential diagnoses are eliminated [6,7]. Electronic diagnoses were confirmed through chart review or standardized examinations.

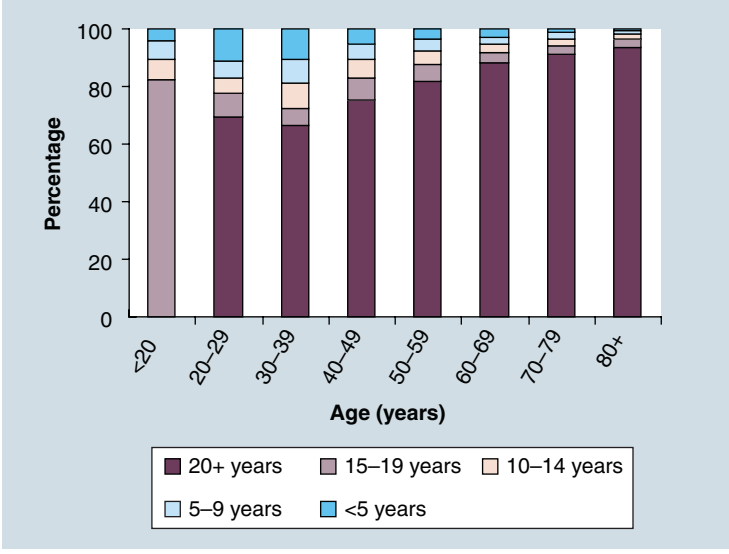
Summary statistics were generated using SAS® 9.1 (SAS Institute, Cary, NC, USA). Specificity and positive predictive values were calculated according to standard methodologies by comparing the 'rule of one' classification with the 'rule of two' electronic classification as the gold standard [8]. nQuery Advisor® Version 4.0 was used to estimate power for hypothetical studies under various assumptions of effect size and risk factor (environment or genotype) prevalence.

Results

A total of 18,239 PMRP participants enrolled from September 2002 to May 2005 were included in the study, with a mean age of 49 years (standard deviation: 18.5), ranging from 18–98 years, of which 57% were female. Most of the cohort had 20 or more years of retrospective medical record data available, while only 6% had less than 5 years of medical record history with the Marshfield Clinic; this varied with age as would be expected (Figure 1). A total of 55% of PMRP subjects are members the Marshfield Clinic-sponsored health maintenance organization, with the percentage covered generally increasing with age (Figure 2).

Of the 18,239 subjects, 86.4% had at least one Healthy People 2010 disease diagnosis. The median number of listed diagnoses per subject was three (range: 1–15). The data in Table 1 demonstrate that the five most common Healthy People 2010 disease conditions in this cohort ('rule of two') are:

Figure 1. Age-specific distribution of years of retrospective clinical data available for the Personalized Medicine Research Project cohort.



- Chronic back conditions (n = 9806; 53.8%)
- Cardiovascular disease including hypertension (n = 7518; 41.2%)
- Obesity (n = 6927; 38.0%)
- Chronic obstructive pulmonary disease (COPD) (n = 6123; 33.6%)
- Arthritis (n = 4814; 26.4%)

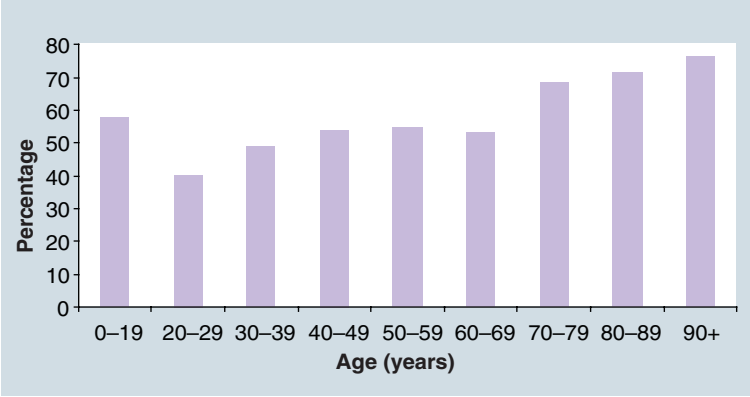
There are fewer than 100 estimated cases of the following Healthy People 2010 diseases: lung cancer (n = 88), angle closure glaucoma (n = 62), oropharyngeal cancer (n = 31), cervical cancer (n = 23) and pneumoconiosis (n = 20). The number of people with birth defects is higher than would be expected, with no obvious explanation.

Specificity values greater than 95% were observed for all of the cancers and 15 other conditions. Specificity values of less than 90% for the electronic algorithm were observed for the following diseases: arthritis (88.0%), cardiovascular disease (85.5%), falls (79.6%), chronic back conditions (75.5%) and unintentional injuries (44.1%). The largest positive predictive value was observed for prostate cancer (91.7%). The smallest positive predictive values were observed for acute, generally nonrecurring events ('E' codes), including unintentional injuries (29.5%), falls (18.9%) and motor vehicle crashes (13.4%).

The median number of annual visits per subject to the Marshfield Clinic was 4.9 (mean: 6.3; range: 0.1–92.9) and the median number of visits in the last year was eight (mean: 13.1; range: 1–167 visits). The median duration of all participants in the Marshfield Clinic system was 29 years (mean: 27.5 years; range: 1 day to 45.3 years). Ranking of the top ten disease conditions from Table 1 by prevalence and by median number of annual visits after diagnosis both yield entirely different lists (Table 2). More than half the population is affected by chronic back conditions but end stage renal disease results in the highest number of median annual visits after diagnosis.

Recognizing that prioritization strategies for the conduct of association studies will typically consider many factors, Figure 3 illustrates the modeling of study power based upon estimated disease prevalence, recognizing that the case status was not validated through chart review or objective examination. With one-to-one matching in a nested case–control study, and assuming 80% power and a conservative minor allele frequency of 0.01, Figure 3 demonstrates the

Figure 2. Age-specific percentage of Personalized Medicine Research Project subjects enrolled in Marshfield Clinic's Security Health Plan.



minimum detectable odds ratios for the 'rule of two' Healthy People 2010 diagnoses in the PMRP cohort. While these estimations suggest that a sample size of at least 1000 cases may be required for the detection of associations with moderate effect size, additional power could be achieved by selecting more than one control per case where possible. In addition, the estimated number of cases of any given disease could decrease with more stringent case definition, for example inclusion of fewer ICD codes.

Conclusions

The Marshfield Clinic PMRP is a population-based resource to facilitate genomic research with the ultimate goal of improving community health. It could also be used for studies of health services utilization, health outcomes and family-based studies because of the self-reported information on family structure and family health history. Policies and procedures have been developed to facilitate access to the database for either internal or external scientists. After scientific merit review and approval from an approved peer review process, all projects must be reviewed and approved at a full board meeting of the Marshfield Clinic Institutional Review Board. Final approval to release the samples is given by an Oversight Committee.

The Healthy People 2010 disease conditions represent one plausible strategy to prioritize research within the PMRP. The data from the present study reveal that the setting of priorities varies according to whether one emphasizes disease prevalence or the impact of a disease on the healthcare system in terms of annual clinic visits, recognizing that annual number of clinic visits is a crude measure of cost to the healthcare system. Although only prevalent disease has been considered for this study, incident cases of disease can be identified because the cohort is being followed longitudinally as they continue to seek care through the Marshfield Clinic system of care.

A limitation of the current project is that disease diagnoses were not confirmed through chart abstraction or objective evaluations of all subjects. However, it is known that the 'rule of two' classification (i.e., requiring a subject to have received the diagnosis on at least two separate visits) improves the specificity for chronic conditions such as Type 2 diabetes. In some cases, we observed large differences in prevalence. For example, the frequency of unintentional injuries, motor vehicle crashes and falls varied considerably when estimated by 'rule of one' rather than 'rule of two' in the current study. This may be due to the fact that many people are seen only once for an acute condition and do not continue to carry that diagnosis in subsequent clinic visits. Prior validation efforts in the PMRP cohort have demonstrated that the use of laboratory values can also improve the accuracy of electronic algorithms to identify cases of disease [7].

Another issue affecting sensitivity and specificity of electronic algorithms to classify disease is the heterogeneity of disease classification. For example, arthritis has several common subclassifications, including osteoarthritis and rheumatoid arthritis. Furthermore, disease can be primary or secondary to another condition or medication use.

The use of disease codes selected for other purposes can help to facilitate standardization and comparison across studies. However, they may not be ideal for all purposes, especially for studies of disease etiology, such as genetic epidemiology studies, where narrowly defined diseases or phenotypes are necessary. An example in the current study is COPD. By 'rule of one', more than half of the PMRP cohort has been diagnosed with COPD. When this finding was investigated further, it was discovered that most diagnoses were due to one code in the range of codes designated as COPD according to the Healthy People 2010 document: 490.0, bronchitis, not specified as acute or chronic. This one code accounted for 8335 (86.3%) of the COPD diagnoses and would fall outside what is usually considered to be chronic obstructive lung disease [9]. By having the data for the individual codes that comprise the large grouping designated in Healthy People 2010, researchers are able to compare their

Table 1. Estimated prevalence of Healthy People 2010 conditions in the Personalized Medicine Research Project cohort.

Healthy People 2010 disease (ordered by HP2010 priority number)	Prevalence (%)		Specificity (%)	Positive predictive value (%)
	Rule 2	*Rule 1*		
Cancer (all sites)	2073 (11.4)	2474 (13.6)	97.5	83.8
Lung cancer	88 (0.5)	114 (0.6)	99.9	77.2
Female breast cancer	368 (3.5)	421 (4.0)	99.5	87.4
Cervical cancer	40 (0.4)	62 (0.6)	99.8	64.5
Colorectal cancer	121 (0.7)	164 (0.9)	99.8	73.8
Oropharyngeal cancer	31 (0.2)	55 (0.3)	99.9	56.4
Prostate cancer	332 (4.3)	362 (4.6)	99.6	91.7
Malignant melanoma	101 (0.6)	119 (0.7)	99.9	84.9
Arthritis	4863 (26.7)	6486 (35.6)	87.9	75.0
Osteoporosis	936 (5.1)	1314 (7.2)	97.8	71.2
Chronic back conditions	9865 (54.1)	11,954 (65.5)	75.1	82.5
Vertebral fractures	241 (1.3)	375 (2.1)	99.3	64.3
Hip fractures	116 (0.6)	200 (1.1)	99.5	58.0
Unintentional injuries	3524 (19.3)	11,917 (65.3)	43.0	29.6
Motor vehicle crashes	285 (1.6)	2158 (11.8)	89.6	13.4
Falls	828 (4.5)	4375 (24.0)	79.6	18.9
Cardiovascular	7548 (41.4)	9135 (50.1)	85.2	82.6
End stage renal disease	207 (1.1)	299 (1.6)	99.5	69.2
Coronary heart disease	2011 (11.0)	2555 (14.0)	96.7	78.7
Stroke	996 (5.5)	1361 (7.5)	97.9	73.2
Diabetes	1734 (9.5)	2168 (11.9)	97.4	80.0
Congenital heart and vascular defects	194 (1.1)	380 (2.1)	99.0	51.1
Chronic obstructive pulmonary disease	6293 (34.5)	9655 (52.9)	71.9	65.2
Pneumoconiosis	517 (2.8)	1171 (6.4)	96.3	44.2
Asthma	2186 (12.0)	2882 (15.8)	95.7	75.9
Cirrhosis	279 (1.5)	484 (2.7)	98.9	57.6
Sleep apnea	699 (3.8)	866 (4.8)	99.1	80.7
Birth defects	2186 (12.0)	4024 (22.1)	88.6	54.3
Peptic ulcer	592 (3.3)	1009 (5.5)	97.6	58.7
Diabetic retinopathy*	329 (19.0)	428 (19.7)	93.2	77.6
Cataract	2499 (13.7)	3150 (17.3)	95.9	79.3
Open angle glaucoma	346 (1.9)	404 (2.2)	99.7	85.6
Angle closure glaucoma	62 (0.3)	91 (0.5)	99.8	68.1
Obesity	6927 (38.2)	6927 (38.2)	NA	NA

Among people with diabetes.

HP2010: Healthy People 2010; NA: Not applicable (because height and weight were measured directly and not abstracted from the medical records).

data, as well as more narrowly define the definition for their specific research purposes. It is important to understand the codes used to electronically identify cases for any given disease.

There is a sizable body of literature on the use of patient record data for research. Several recent reviews of the quality of information in electronic medical records reveal variability between systems [10–12], primarily due to different disease definitions, and variability within systems, due to the distinctiveness of disease diagnoses. Despite the limitations observed in using electronic medical records to classify disease state, the electronic estimates of disease

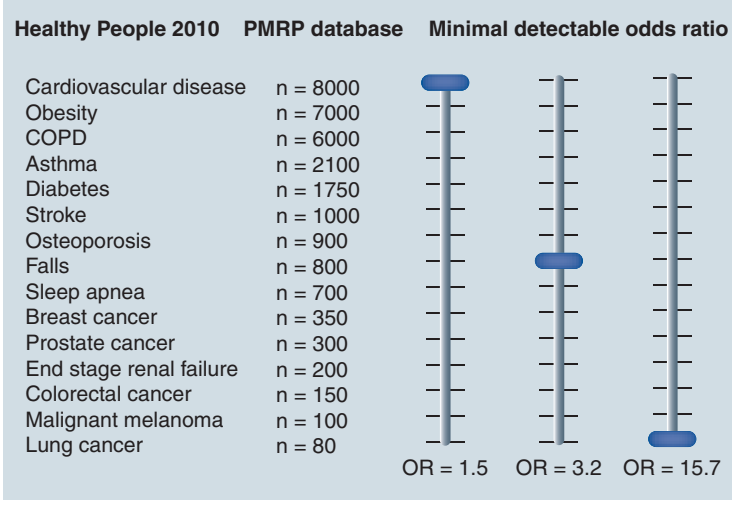
Table 2. Rank of disease conditions in the Personalized Medicine Research Project cohort by disease prevalence and median number of annual visits after diagnosis.

Rank	Disease prevalence	Median number of annual visits after diagnosis
1	Chronic back conditions (54%)	End stage renal disease (29 visits)
2	Cardiovascular (41%)	Lung cancer (25 visits)
3	Obesity (38%)	Oropharyngeal cancer (20 visits)
4	Chronic obstructive pulmonary disease (34%)	Sleep apnea (18 visits)
5	Arthritis (27%)	Diabetic retinopathy (18 visits)
6	Unintentional injuries (19%)	Stroke (18 visits)
7	Cataract (14%)	Hip fractures (18 visits)
8	Asthma (12%)	Coronary heart disease (17 visits)
9	Birth defects (12%)	Colorectal cancer (16 visits)
10	Cancer, all sites (11%)	Prostate cancer (16 visits)

prevalence obtained within this study cohort appear to provide a good starting point for planning and prioritization of research ideas.

It has been suggested that common genetic variants in the population increase susceptibility to common diseases and that the effect sizes for new discoveries are likely to be relatively small for any single gene [13]. A representative power calculation reveals that the PMRP cohort will have over 90% to detect an odds ratio as small as 1.75 for diseases at least as common as 5% (osteoporosis, number 15 on the list of Healthy People 2010 diseases). For a genome-wide association approach [2], new statistical approaches, such as multifactor dimensionality reduction, are being developed to account for the potential problem of multiple comparisons in cohorts such as the PMRP [14].

Figure 3. Prevalence-driven power modeling in a large population-based DNA biobank.



COPD: Chronic obstructive pulmonary disease; OR: Odds ratio; PRMP: Personalized Medicine Research Project.

Another strategy for research prioritization is community-based participatory research [15,16], and the PMRP is engaged in this effort through its Community Advisory Group that meets twice a year [4]. There is huge potential to expand in this area.

Recruitment into the PMRP is ongoing, with a target of approximately 20,000 participants by 2007. After attaining a population-based cohort of approximately 20,000 subjects, additional recruitment will be targeted towards disease cohorts for the less common conditions, such as the specific cancers, with a target of 500–1000 cases of any specific disease to be studied. For these diseases, targeted recruitment to facilitate nested case–control studies is a more efficient study design than continued population-based recruitment. These data will also help to direct future targeted, disease-specific recruitment to augment the existing population-based cohort. The selection of population-based controls for the additional nested cases targeted from throughout the clinic system could prove to be a valid variation on the initial population-based design. Controls can be selected electronically, depending on the condition of interest and whether there is a recommended screening for the disease. For example, breast cancer cases can be identified electronically through the tumor registry. Valid controls can be selected by identifying women who have never had a diagnosis of breast cancer and who have had a mammogram within a certain period of time. The same would hold true for hypercholesterolemia or diabetes, where there are standard screening recommendations and available electronic data for blood cholesterol and blood sugar measurements. For diseases where there are no routine screening guidelines, such as asthma, the potential to

misclassify controls is greater and may require that the controls be screened specifically. This procedure was used within the PMRP in a study of Alzheimer's disease. Cases were identified purely on the basis of medical record abstraction and potential controls were screened through telephone administration of a Mini-Mental State Examination. Upon initial enrollment, more than 99% of the PMRP cohort agreed to contact for future studies.

In summary, the PMRP cohort is currently large enough to study most diseases of public health importance. The Healthy People 2010 document provides a prioritization strategy that can be applied to diseases within this cohort and the PMRP cohort can be used to track progress towards the 2010 goals in a rural setting. At present, the PMRP database is available for use by scientists at the Marshfield Clinic and external scientists.

Future perspective

Many research groups around the world are developing biobanks for genetic and genomic research, with the ultimate aim of identifying genetic markers to improve prediction, diagnosis and treatment: 'personalized medicine'. To decrease the time from discovery to clinical translation, these groups will need to work together to study rare diseases and this will require standardization of research tools. Prospective studies in clinical settings must be conducted and must include an assessment of cost so that third party payers recognize the value of predictive genetic testing and thus would be willing to pay for it. The next 5–10 years will see the community at large clamoring for the timely translation of research discoveries and this will be supported through priorities for funding at all levels and may be further pushed along through potential litigation.

Executive summary

- The Personalized Medicine Research Project (PMRP) is a population-based biobank with more than 18,000 adults aged 18–98 years enrolled. The purpose of the PMRP is to provide a resource to facilitate research in the areas of genetic epidemiology, pharmacogenetics and population genetics. Issues being considered include:
  - How to prioritize research within the biobank given finite resources. Disease prevalence or impact on the healthcare system in terms of annual number of Clinic visits are two possibilities. Community-based participatory research is a growing strategy for the prioritization and implementation of research.
  - After a target of 20,000 population-based subjects, disease-specific recruitment will commence within the PMRP, with a target of at least 500–1000 cases of any specific disease to allow genome wide-association studies.
  - Electronic algorithms need to be developed and validated to make use of the vast amount of data contained in electronic medical records and to allow comparisons across studies.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. *Genetics and Public Health in the 21st Century.* Khoury MJ, Burke W, Thompson EJ (Eds). Oxford University Press, Oxford, UK (2000).
2. Christensen K, Murray JC: What genome-wide association studies can do for medicine. *N. Engl. J. Med.* 356, 1094–1097 (2007).

• Nice overview of genome-wide association studies with a simple explanation of genetic concepts.

3. Collins FS, Green ED, Guttmacher AE, Guyer MS; US National Human Genome Research Institute: A vision for the future of genomics research. A blueprint for the genomic era. *Nature* 422, 835–847 (2003).

•• Published to coincide with the completion of the Human Genome Project. Outlines the future directions for genomic research as envisioned by the leaders of the National Human Genome Research Institute.

4. McCarty CA, Wilke RA, Giampietro PF, Wesbrook SD, Caldwell MD: Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a population-based biobank. *Personalized Med.* 2, 49–79 (2005).
5. Wisconsin Turning Point Transformation Team: *Healthiest Wisconsin 2010: A Partnership Plan to Improve the Health of the Public.* Wisconsin Department of Health and Family Services Division of Public Health, Madison, WI, USA.
6. DeStefano F, Eaker ED, Broste SK *et al.*: Epidemiologic research in an integrated regional medical care system: the Marshfield Epidemiologic Study Area. *J. Clin. Epidemiol.* 49, 643–652 (1996).
7. Wilke RA, Berg RL, Peissig P *et al.*: Use of an electronic medical record for the identification of research subjects with type II diabetes. *Clin. Med. Res.* (2007) (In Press).
8. Riegelman RK, Hirsch RP: *Studying a Study and Testing a Test. How to Read the Medical Literature*. Little, Brown and Company, Boston, MA, USA (1989).
9. Soriano JB, Davis KJ, Coleman B, Visick G, Mannino D, Pride NB: The proportional venn diagram of obstructive lung disease. Two approximations from the United States and the United Kingdom. *Chest* 124, 474–481 (2003).
10. Thiru K, Hassey A, Sullivan F: Systematic review of scope and quality of electronic patient record data in primary care. *Br. Med. J.* 326, 1070–1074 (2003).
11. Jordan K, Porcheret M, Croft P: Quality of morbidity coding in general practice computerized medical records: a systematic review. *Fam. Pract.* 21, 396–412 (2004).
12. Schneeweiss S, Avorn J: A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J. Clin. Epidemiol.* 58, 323–337 (2005).

13. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* 33, 177–182 (2003).
• Nice overview of genetic association studies and some of the common problems of false-negative results due to underpowered small studies, thus supporting the need for larger studies.
14. Hahn LW, Ritchie MD, Moore JH: Multifactor dimensionality reduction software for detecting gene–gene and gene–environment interactions. *Bioinformatics* 19, 376–382 (2003).
15. Martin D, Abelson J, Singer P: Participation in health care priority-setting through the eyes of the participants. *J. Health Serv. Res. Policy* 7, 222–229 (2002).
16. RTI International University of North Carolina: *Community-Based Participatory Research: Assessing the Evidence.* Agency for Healthcare Research and Quality, Rockville, MD, USA, Publication No. 04-E02202 (2004).
•• Excellent summary of the state of knowledge and research regarding community-based participatory research. Dr Zerhouni, director of the US NIH, has stated that 'participatory' is one of the four P's in the future paradigm for translational research.

Websites

101. Centers for Disease Prevention and Control – National Office of Public Health Genomics
www.cdc.gov/genomics
102. Marshfield Clinic Research Foundation – Personalized Medicine Research Project
www.marshfieldclinic.org/chg/pages/default.aspx?page=chg_pers_med_res_prj
103. Office of Disease Prevention and Health Promotion, US Department of Health and Human Services – Healthy People 2010
www.healthypeople.gov/
104. Marshfield Clinic
www.marshfieldclinic.org