Appendix 1: Research Protocol

Systematic review of the efficacy of different methods for the diagnosis of hypertension

Background

There is no uniform consensus in the frequency and timing of measurements of BP for diagnosis of hypertension. The NICE and British Hypertension Society Guidelines recommend measurements over 3-6 months and a threshold of sustained blood pressure in excess of 160/100 mmHg (uncomplicated) or 140/90mmHg (in the presence of additional risk factors).[1,2] European and US clinical guidelines for self-monitoring state the initial assessment of BP should be for a seven day period with two recordings on each occasion in the morning and the evening and the first days readings discarded before taking a mean.[3,4] A threshold of 135/85 mmHg is suggested but systematic reviews and other international guidelines vary in the threshold used depending on the method of ascertainment. Ambulatory monitoring can give an answer after 24 hours of measurements and a threshold here of 135/85 mmHg mean day time BP is quoted.[5] (Appendix A details the diagnostic thresholds for hypertension recommended across a number of different national and international guidelines for clinic/ office, home and ambulatory blood pressure from to the current day.) This study will ascertain the current literature regarding the relative effectiveness of these methods in the diagnosis of hypertension as well as white coat hypertension and masked hypertension.

Objective

To ascertain from the worldwide literature the performance characteristics of clinic measurments, home blood pressure monitoring and ambulatory monitoring in the diagnosis of hypertension.

Research question:
What are the performance characteristics of the following methods of blood pressure measurement in the diagnosis of hypertension in terms of the relative proportions diagnosed (or not) by each of (a) clinic measurements, (b) home blood pressure monitoring and (c) ambulatory monitoring, as compared to each other?

Criteria for considering studies for this review

Types of studies:
Only studies comparing at least two methods of blood pressure measurement in the diagnosis of hypertension and with extractable data to populate 2x2 tables will be included. In addition, if there is no clear definition of how the method of monitoring was applied or no clear information provided on the cut-off point for diagnosis of hypertension, studies will be excluded. Studies will not be excluded based on year.

Types of participants in studies:
Patients will be excluded if they are pregnant or hospitalised. They will also be excluded if hypertension was already diagnosed or they were on treatment for hypertension. Adult patients of all ages will be included, but not children.

Types of interventions:
Any study comparing at least two of the methods of blood pressure measurement in the diagnosis of hypertension may be included. Studies obtaining their population using screening are eligible for inclusion. Studies looking only at outcome against method of blood pressure measurement will not be included.

Types of outcome measures:
The main output will be 2x2 tables for each comparison of the measurement methods. Studies will only be included if measurements are provided for at least two of the three methods of monitoring (i.e. clinic, home and ambulatory) and there is extractable data to populate 2x2 tables.

Search strategy for identification of studies

Medline, Embase, the Cochrane database of systematic reviews, DARE (to identify anything new since the last Cochrane update), Medion (http://www.mediondatabase.nl), ARIF (http://www.arif.bham.ac.uk and the TRIP database will be searched for articles published up to and including to end 2008, using a search strategy designed to capture all studies evaluating the test performance characteristics of different methods of diagnosing hypertension in primary care. No language or publication date limits will be applied.

The search strategy is based on the diagnostic filters developed by Haynes et al[6] and Montori et al.[7] However, recognising that these filters rely on indexing terms for research methodology and text words used in reporting results[8], to improve sensitivity in the search three separate strategies will be combined when using Medline and Embase:
1: Combining keywords for hypertension, blood pressure monitoring, outpatient setting and diagnosis
2: Limiting MeSH terms for hypertension to diagnosis sub heading, and combining this with keywords for blood pressure monitoring and outpatient setting
3: Combining keywords for hypertension, blood pressure monitoring, outpatient setting and limit using the diagnosis search filter

Medline search strategy:
1. hypertension/ or hypertension, malignant/ or exp hypertension, renal/
2. hypertens*.tw.
3. 1 or 2
4. Blood Pressure/
5. exp Blood Pressure Determination/
6. exp Sphygmomanometers/
7. (sphygmomanometer* or (blood pressure adj3 (monitor* or determin*))).tw.
8. 6 or 5 or 4 or 7
9. Ambulatory Care/
10. ambulatory care facilities/ or outpatient clinics, hospital/
11. Family Practice/
12. Monitoring, Ambulatory/
13. Physicians, Family/
14. "Referral and Consultation"/
15. (family practi* or general practi* or family physician* or primary care).tw.
16. (ambula* adj5 care).tw.
17. (referral* adj5 consultat*).tw.
18. 13 or 9 or 11 or 14 or 17 or 16 or 10 or 12 or 15
19. 8 and 3
20. 8 and 3 and 18
21. diagnos*.tw.
22. (physical exam* or clinical exam*).tw.
23. stress test*.tw.
24. (sensitiv* or specific* or predictive value* or ppv or npv or likelihood ratio* or interobserver or intraobserver).tw.
25. (accuracy or precision or reliability or validity).tw.
26. "Diagnostic Techniques and Procedures"/
27. exp Diagnostic Errors/
28. Diagnostic Tests, Routine/
29. exp "Sensitivity and Specificity"/
30. diagnosis/ or diagnosis, differential/
31. 25 or 23 or 30 or 26 or 24 or 22 or 28 or 21 or 29 or 27
32. 31 and 20
33. Hypertension/di [Diagnosis]
34. Hypertension, Malignant/di [Diagnosis]
35. 33 or 34
36. 8 and 18
37. 35 and 36
38. limit 20 to "diagnosis (sensitivity)"
39. 38 or 32 or 37

The searches on Trip, Medion and Arif will be conducted by looking for "blood pressure" in the titles.

Additionally, reference lists from included studies and previous meta-analyses will be searched. Experts in blood pressure diagnosis will be contacted to identify grey literature not already captured.

Methods of the review

Selection of trials
The search strategy will identify all relevant articles and these will be reviewed in three stages. Two reviewers (RM, JH) will independently review firstly the titles and secondly the abstracts of the articles for potential relevance to the research question. A study will be eligible for the next stage if either reviewer includes it. Finally, those articles remaining will be called and the full papers assessed according to the following criteria (by JH and RM):
1. Does the paper compare more than one method of BP measurement in the diagnosis of hypertension?
2. Is the comparison data extractable? (into a 2x2 table format)
3. Are patients off hypertension treatment at the time of being assessed for a diagnosis of hypertension?

If the answer is negative to any of these three questions, the study will be excluded.

A fourth criterion will also be applied at the stage of full papers being received, namely: Was the threshold for diagnosis used in clinic/ office blood pressure monitoring 140/90 mmHg? Those studies that did use this diagnostic threshold will be considered separately (and initially) to those using some other threshold.

Data management and extraction
Data extraction will be carried out by four reviewers (JH, RM, UM, JM). Initial data extraction on the details of the paper itself (authors, year), the population of the study (country, age, gender, ethnicity profile), sample size and drop-out rate and the setting for the intervention will be done by one reviewer (JH) and checked by one other reviewer separately.

All other data will be extracted independently by two of the four reviewers. Information will be collected on the inclusion and exclusion criteria for the study, the methodological design of the study, the particular methods of blood pressure monitoring being compared, the threshold(s) used to diagnose hypertension, whether risk was taken into account in deciding the threshold, whether any sensitivity analysis using different thresholds was carried out, the details of the monitoring – number and frequency of measurements, definition of monitoring used, the type of monitor used, the order of the measurements, time of day of measurement, which arm used, and the personnel who carried out the monitoring – and the results of the studies (into 2x2 tables of comparisons and mean blood pressure by measurement method if available). Differences in data extraction will be resolved by referring back to the original article and establishing consensus. A third reviewer will be consulted to resolve any remaining differences.

When necessary the authors of the primary studies will be contacted to obtain additional information.

A full copy of the data extraction form is in Appendix B. Data will be managed using an electronic database / spreadsheet (it will be extracted direct into Excel or onto a copy of this form).

Assessment of methodological quality
As part of the data extraction, information will be collected on recognised sources of bias in diagnostic test accuracy studies using a version of the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) checklist[9] adapted for this study (it is recognised that an overall quality *score* should not be used because different shortcomings may generate distinct magnitudes of bias, even in opposing directions[8]). We will consider spectrum bias and selection bias in how the sample was formed (i.e. everyone on the practice list or a specific subset – are all patients having a diagnosis of hypertension considered?), and how were the patients chosen (e.g. consecutive, random or otherwise), sample size, rate of drop-out, were the different monitoring methods conducted independently of each other to detect any partial verification or differential verification bias, information bias in whether

investigators and patients were blinded (i.e. was the result of previous monitoring concealed or not), and whether and how readings of self-monitoring were checked (patient-recorded or telemonitoring).

Data synthesis
We will firstly follow the standard diagnostic accuracy paradigm, though this assumes one of the approaches is the perfect or reference standard. We will provide a meta-analysis of 2x2 tables results for each comparison, in this case necessarily focusing on a binary classification of data rather than using continuous data, and then consider the discrepant cells. We will calculate Kappa to ascertain how much the tests disagree as well as how often. We will extract paired estimates of test sensitivity and specificity from each study overall and plot the studies on summary ROC (receiver-operating characteristic) curve plots. For the subset of studies where the combined data share a common threshold (clinic/ office blood pressure monitoring 140/90 mmHg) we will compute average values of sensitivity and specificity. We will use a hierarchical summary ROC or bivariate random-effects model to account for sampling variability, unexplained heterogeneity and covariation between sensitivity and specificity[8], to enable exploration of heterogeneity and the effect of different diagnostic thresholds.

We will conduct sensitivity analyses to consider the effect of: (a) differing the diagnostic thresholds; (b) population characteristics, including assessing test performance in populations with mean clinic blood pressure at or above the diagnostic threshold, in order to separately consider where study populations have been recruited entirely from a typical patient screening population (and so excluding any studies where an additional group of normotensives were included as 'controls'); and (c) variability in study methodological quality and monitoring methodology.

Interpretation
We will consider the consequences of using the different measurement approaches in terms of the numbers of positive and negative results in light of the expected prevalence of hypertension. We will address the applicability of the results in terms of whether the tests evaluated and compared in the primary studies were representative of those used in practice, and also to what extent the original studies were biased and how these biases could influence the results and the degree to which comparisons between tests may be confounded.

Potential conflict of interest
All co-investigators declare that they have no conflict of interest.

Sources of support
This work forms part of a larger programme considering monitoring in primary care and supported by the NIHR.

[1] CG34 Hypertension (persistently high blood pressure) in adults - NICE guideline. National Institute for Clinical Excellence. http://www.nice.org.uk/guidance/index.jsp?action=byID&o=10986. Accessed 12-8-0008.
[2] Williams B, Poulter NR, Brown MJ, Davis M, McInnes GT, Potter JF, Sever PS, McG TS. Guidelines for management of hypertension: report of the fourth working party of the British Hypertension Society, 2004-BHS IV. *J Hum.Hypertens* 2004;**18**:139-185.
[3] Pickering TG, Miller NH, Ogedegbe G, Krakoff LR, Artinian NT, Goff D. Call to action on use and reimbursement for home blood pressure monitoring: a joint scientific statement from the American Heart Association, American Society Of Hypertension, and Preventive Cardiovascular Nurses Association. *Hypertension* 2008;**52**:10-29.
[4] Parati G, Stergiou GS, Asmar R, et al. European Society of Hypertension guidelines for blood pressure monitoring at home: a summary report of the Second International Consensus Conference on Home Blood Pressure Monitoring. *J Hypertens.* 2008;**26**:1505-1526.
[5] O'Brien E, Asmar R, Beilin L, et al. European Society of Hypertension recommendations for conventional, ambulatory and home blood pressure measurement. *J Hypertens.* 2003;**21**:821-848.
[6] Haynes RB, Wilczynski NC for the Hedges Team. Optimal search strategies for retrieving scientifically strong studies of diagnosis from MEDLINE: analytical survey. *BMJ*. 2004 May 1;328(7447):1040. Epub 2004 Apr 08. PMID: 15073027
[7] Montori VM, Wilczynski NL, Morgan D, Haynes RB; Hedges Team. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ*. 2005 Jan 8;330(7482):68. Epub 2004 Dec 24. PMID: 15619601

[8] Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM on behalf of the Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008; 149: 889-897.

[9] Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003; 3: 25. PMID: 14606960.